The Astrophysical Journal, 784:143 (21pp), 2014 April 1 doi:10.1088/0004-637X/784/2/143

A MEASUREMENT OF THE SECONDARY-CMB AND MILLIMETER-WAVE-FOREGROUND BISPECTRUM USING 800 deg^2 OF SOUTH POLE TELESCOPE DATA

T. M. Crawford[1,2], K. K. Schaffer[1,3,4], S. Bhattacharya[1,5], K. A. Aird[6], B. A. Benson[1,3], L. E. Bleem[1,7], J. E. Carlstrom[1,2,3,7,8], C. L. Chang[1,3,8], H-M. Cho[9], A. T. Crites[1,2], T. de Haan[10], M. A. Dobbs[10], J. Dudley[10], E. M. George[11], N. W. Halverson[12], G. P. Holder[10], W. L. Holzapfel[11], S. Hoover[1,7], Z. Hou[13], J. D. Hrubes[6], R. Keisler[1,7], L. Knox[13], A. T. Lee[11,14], E. M. Leitch[1,2], M. Lueker[15], D. Luong-Van[6], J. J. McMahon[16], J. Mehl[1,8], S. S. Meyer[1,2,3,7], M. Millea[13], L. M. Mocanu[1,2], J. J. Mohr[17,18,19], T. E. Montroy[20], S. Padin[1,2,15], T. Plagge[1,2], C. Pryke[21], C. L. Reichardt[11], J. E. Ruhl[20], J. T. Sayre[20], L. Shaw[10], E. Shirokoff[11], H. G. Spieler[14], Z. Staniszewski[20], A. A. Stark[22], K. T. Story[1,7], A. van Engelen[10], K. Vanderlinde[10], J. D. Vieira[15], R. Williamson[1,2], and O. Zahn[23]

[1] Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA; tcrawfor@kicp.uchicago.edu
[2] Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
[3] Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA
[4] Liberal Arts Department, School of the Art Institute of Chicago, Chicago, IL 60603, USA
[5] High Energy Physics Division, Argonne National Laboratory, Argonne, IL 60439, USA
[6] University of Chicago, Chicago, IL 60637, USA
[7] Department of Physics, University of Chicago, Chicago, IL 60637, USA
[8] Argonne National Laboratory, Argonne, IL 60439, USA
[9] NIST Quantum Devices Group, Boulder, CO 80305, USA
[10] Department of Physics, McGill University, Montreal, Quebec H3A 2T8, Canada
[11] Department of Physics, University of California, Berkeley, CA 94720, USA
[12] Department of Astrophysical and Planetary Sciences and Department of Physics, University of Colorado, Boulder, CO 80309, USA
[13] Department of Physics, University of California, Davis, CA 95616, USA
[14] Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[15] California Institute of Technology, Pasadena, CA 91125, USA
[16] Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[17] Department of Physics, Ludwig-Maximilians-Universität, D-81679 München, Germany
[18] Excellence Cluster Universe, D-85748 Garching, Germany
[19] Max-Planck-Institut für extraterrestrische Physik, D-85748 Garching, Germany
[20] Physics Department, Center for Education and Research in Cosmology and Astrophysics, Case Western Reserve University,Cleveland, OH 44106, USA
[21] Department of Physics, University of Minnesota, Minneapolis, MN 55455, USA
[22] Harvard-Smithsonian Center for Astrophysics, Cambridge, MA 02138, USA
[23] Berkeley Center for Cosmological Physics, Department of Physics, University of California, and Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Received 2013 March 13; accepted 2014 February 24; published 2014 March 14

ABSTRACT

We present a measurement of the angular bispectrum of the millimeter-wave sky in observing bands centered at roughly 95, 150, and 220 GHz, on angular scales of $1' \lesssim \theta \lesssim 10'$ (multipole number $1000 \lesssim l \lesssim 10{,}000$). At these frequencies and angular scales, the main contributions to the bispectrum are expected to be the thermal Sunyaev–Zel'dovich (tSZ) effect and emission from extragalactic sources, predominantly dusty, star-forming galaxies (DSFGs) and active galactic nuclei. We measure the bispectrum in 800 deg^2 of three-band South Pole Telescope data, and we use a multi-frequency fitting procedure to separate the bispectrum of the tSZ effect from the extragalactic source contribution. We simultaneously detect the bispectrum of the tSZ effect at $>10\sigma$, the unclustered component of the extragalactic source bispectrum at $>5\sigma$ in each frequency band, and the bispectrum due to the clustering of DSFGs—i.e., the clustered cosmic infrared background (CIB) bispectrum—at $>5\sigma$. This is the first reported detection of the clustered CIB bispectrum. We use the measured tSZ bispectrum amplitude, compared to model predictions, to constrain the normalization of the matter power spectrum to be $\sigma_8 = 0.787 \pm 0.031$ and to predict the amplitude of the tSZ power spectrum at $l = 3000$. This prediction improves our ability to separate the thermal and kinematic contributions to the total SZ power spectrum. The addition of bispectrum data improves our constraint on the tSZ power spectrum amplitude by a factor of two compared to power spectrum measurements alone and demonstrates a preference for a nonzero kinematic SZ (kSZ) power spectrum, with a derived constraint on the kSZ amplitude at $l = 3000$ of $A_{\rm kSZ} = 2.9 \pm 1.6\,\mu{\rm K}^2$, or $A_{\rm kSZ} = 2.6 \pm 1.8\,\mu{\rm K}^2$ if the default $A_{\rm kSZ} > 0$ prior is removed.

Key words: cosmic background radiation – cosmology: observations – methods: data analysis

Online-only material: color figures

1. INTRODUCTION

Measurements of the millimeter-wave sky are a rich source of cosmological information. Studies of the intensity of the cosmic microwave background (CMB) have provided much of the evidence for our current cosmological model (e.g., Komatsu et al. 2011; Hinshaw et al. 2013), and ever more sensitive and wide-ranging experiments (in terms of both sky coverage and range of angular scales probed) continue to improve our constraints on cosmological parameters (Reichardt et al. 2009; Das et al. 2011b; Keisler et al. 2011; Story et al. 2013; Hou et al. 2014; Sievers et al. 2013). Beyond revealing the properties of the primary CMB fluctuations (those generated at the surface of last scattering), the high-resolution millimeter-wave sky maps

generated by current experiments also enable the study of secondary anisotropies in the CMB—those due to interactions of the CMB photons with matter along the line of sight—as well as emissive foreground sources. These secondary signals also carry interesting cosmological information, probing different epochs of cosmic history (e.g., Lueker et al. 2010; Das et al. 2011b; Reichardt et al. 2012).

The primary CMB signal arising from fluctuations at the surface of last scattering is expected to be very close to a Gaussian random field. The simplest models of inflation predict very small levels of non-Gaussianity (e.g., Acquaviva et al. 2003), and experimental results up to this point have been essentially consistent with these predictions (Komatsu et al. 2009, 2011). Under the assumption of Gaussianity, all of the information about CMB intensity fluctuations is contained in the second moment of the distribution. Hence, the power spectrum has historically been the simplest and most useful statistic for characterizing CMB fluctuations and constraining cosmological models.

Searches for non-Gaussianity in the primary CMB, as pointed out in, e.g., Coles & Barrow (1987), have the potential to inform inflationary models. These measurements involve constructing statistics that test the skewness of the maps, or, more generally, that depend on the three-point angular correlation function or its harmonic-space equivalent, the bispectrum. Such statistics can be tailored for sensitivity to specific models for inflationary non-Gaussianity, which are then parameterized with a single amplitude $f_{\rm NL}$. Several analyses of *Wilkinson Microwave Anisotropy Probe* (*WMAP*) data have produced conflicting constraints on the amplitude of a particular model—the so-called local model—ranging from tentative detections to limits consistent with $f_{\rm NL,local} = 0$, but always with error bars of $\delta f_{\rm NL} \sim 20$–$30$ (e.g., Yadav & Wandelt 2008; Komatsu et al. 2011). Analyses of higher-resolution CMB data over small patches of the sky have resulted in upper limits of order $f_{\rm NL,local} < 1000$ (e.g., Santos et al. 2002; Smith et al. 2004). More precise constraints or measurements of $f_{\rm NL}$ will require a combination of large sky area and high sensitivity; limits from *Planck* are expected to be at the level of $\delta f_{\rm NL} \sim 5$ (e.g., Yadav et al. 2007). Progress will also depend on understanding the non-Gaussian behavior of secondary CMB anisotropies and foregrounds.

Secondary CMB anisotropies include gravitational lensing of the CMB (e.g., Hu 2000) and the thermal Sunyaev–Zel'dovich (tSZ) effect (Sunyaev & Zel'dovich 1970). These signals are non-Gaussian in general, although they are often studied through their effect on the power spectrum. For example, detections of the effect of gravitational lensing in CMB data have often relied on the effect of acoustic peak smearing in the two-point function or power spectrum (e.g., Reichardt et al. 2009; Das et al. 2011b, 2013; Keisler et al. 2011; Story et al. 2013), but more information can be extracted from CMB lensing when the four-point function is considered (e.g., Das et al. 2011a; van Engelen et al. 2012).

The tSZ effect arises from the spectral distortion of the CMB through interactions with hot gas in galaxy clusters and provides an efficient means of finding distant, massive clusters (e.g., Williamson et al. 2011; Planck Collaboration et al. 2011; Reichardt et al. 2013; Hasselfield et al. 2013). The tSZ power spectrum measures the mean squared signal from all clusters and is a sensitive probe of the normalization of the matter power spectrum σ_8, with the tSZ power spectrum amplitude predicted to scale as σ_8^{γ}, with $\gamma \sim 7$–8 (Komatsu & Seljak 2002; Reichardt et al. 2012). The tSZ power spectrum amplitude at few-arcminute scales ($l \sim 3000$) has been constrained through recent measurements of the small-angular-scale power spectrum of the millimeter-wave sky (e.g., Lueker et al. 2010; Das et al. 2011b; Reichardt et al. 2012). However, the tSZ power spectrum at these scales is dominated by high-redshift, low-mass clusters that are not well studied at other wavelengths (e.g., Holder 2002). Significant modeling uncertainty for this population of clusters complicates the translation of the measured tSZ power spectrum amplitude into a constraint on σ_8 (e.g., Lueker et al. 2010).

An alternative approach to constraining σ_8 and other parameters that affect the growth of structure is to study just those galaxy clusters that can be individually detected in millimeter-wave maps through their tSZ signature. When redshifts are obtained for every cluster, this approach can constrain both σ_8 and the equation of state of dark energy (e.g., Wang & Steinhardt 1998; Haiman et al. 2001). The scaling of the observable (the number of clusters detected) with σ_8 is even steeper than for the power spectrum, with number counts going roughly as σ_8^{10} (e.g., Dudley 2012). Constraints based on number counts are nearly independent of those using the tSZ power spectrum, making the two probes nicely complementary.

The thermal SZ bispectrum offers another approach that complements both the power spectrum and cluster-detection methods. As shown in Bhattacharya et al. (2012, hereafter B12), the $l \sim 3000$ tSZ bispectrum is dominated by more massive, lower-redshift clusters than the tSZ power spectrum at similar angular scales. This population of clusters is subject to less modeling uncertainty than the higher-redshift, lower-mass clusters that dominate the tSZ power spectrum. Furthermore, the tSZ bispectrum is yet more sensitive to σ_8 than the tSZ power spectrum and cluster number counts. B12 demonstrated that the amplitude of the tSZ bispectrum at $l = 3000$ scales as $B_{\rm tSZ} \propto \sigma_8^{11-12}$. Hill & Sherwin (2013) demonstrated a similar scaling for the real-space tSZ skewness $\langle T_{\rm tSZ}^3 \rangle$, and Wilson et al. (2012) used this model prediction and a measurement of the tSZ skewness in Atacama Cosmology Telescope (ACT) data to place a $\lesssim$5% constraint on σ_8.

B12 also showed that by simultaneously constraining cosmology and cluster physics with the tSZ bispectrum, one could make a precise prediction for the amplitude of the tSZ power spectrum. A measurement of the tSZ bispectrum provides new constraints on intracluster gas physics and therefore acts as a bridge between the very low redshift, very massive clusters that currently constrain gas models (mostly through X-ray observations), and the very high redshift, low-mass clusters that dominate the tSZ power spectrum.

The tSZ contribution to the CMB power spectrum is difficult to separate from the contribution of the kinematic SZ (kSZ) effect (Sunyaev & Zel'dovich 1980) in current data (Reichardt et al. 2012). However, under the assumption that the contribution to the bispectrum from the kSZ is negligible (see Section 4.1 for details), the bispectrum-based prediction for the tSZ power spectrum can be used to sharpen the measurement of the kSZ power spectrum. The resulting constraints on the kSZ effect—which are interesting for cosmology and for models of reionization (e.g., Knox et al. 1998; Gruzinov & Hu 1998)—are potentially much stronger than from the power spectrum alone.

Finally, the non-Gaussian signals from extragalactic emissive sources are also potentially interesting. Two populations of sources contribute significantly to measurements at the frequencies at which CMB experiments typically operate (roughly

a few GHz to hundreds of GHz): synchrotron-dominated "radio sources"—primarily active galactic nuclei—and dusty star-forming galaxies (DSFGs), the integrated light from which produces the cosmic infrared background (CIB). Measurements of the bispectrum contribution from radio sources and DSFGs can constrain source count models beyond the threshold for individually detecting sources, and with different flux weighting than measurements using the power spectrum. Perhaps more intriguingly, measurements of the bispectrum of the clustered CIB (fluctuations in the mean CIB emission due to large-scale structure) have the potential to constrain models of galaxy and star formation beyond what can be done with CIB power spectrum measurements.

The bispectra of the secondary anisotropies and foregrounds are characterized by different angular scale dependence as well as different spectral signatures. Multi-band measurements across a wide range of angular scales can therefore be used to isolate the different contributions. The purpose of this work is to present bispectrum measurements in three frequency bands centered at roughly 95, 150, and 220 GHz, (corresponding to wavelengths of ∼3.2, ∼2.0, and ∼1.4 mm), using approximately 800 deg^2 of sky from the South Pole Telescope SZ (SPT-SZ) survey. We concentrate on the range of angular scales (or multipole number) at which the secondary and foreground sources of non-Gaussianity are expected to dominate, namely $\theta \lesssim 10'$ ($l \gtrsim 1000$), and use this information to simultaneously fit for the contributions from tSZ and from the clustered and spatially uncorrelated contributions from emissive sources.

The paper is organized as follows. We briefly describe the data products used in this analysis in Section 2, we present the method for estimating the bispectrum in Section 3, we describe the model used to fit the resulting bispectrum measurements in Section 4, we present measured bispectra and model fit results in Section 5, we discuss the implications of these results for cosmology and models of source emission in Section 6, and we conclude in Section 7.

2. DATA

The SPT (Carlstrom et al. 2011) is a 10 m telescope located at the National Science Foundation's Amundsen-Scott South Pole station in Antarctica. The 2500 deg^2 SPT-SZ survey was completed in 2011 November. This survey produced maps in three frequency bands (95, 150, and 220 GHz) to a depth such that the rms 150 GHz noise level in any of the maps is $<18\,\mu$K per 1 arcmin pixel. Scientific results from partial or full SPT-SZ survey data include catalogs of clusters discovered via the SZ effect (Vanderlinde et al. 2010; Williamson et al. 2011; Reichardt et al. 2013), catalogs of emissive sources (including a new population of strongly lensed, dusty, high-redshift galaxies, Vieira et al. 2010, 2013), and measurements of the primary CMB power spectrum (Keisler et al. 2011; Story et al. 2013) and of the secondary CMB and foreground power spectra (Lueker et al. 2010; Shirokoff et al. 2011; Reichardt et al. 2012).

Reichardt et al. (2012, hereafter R12) used ∼800 deg^2 of SPT-SZ survey data to measure the small-angular-scale CMB power spectrum and place unprecedentedly tight constraints on the fluctuation power of the tSZ, kSZ, and CIB at SPT observing frequencies. In this work, we measure the bispectrum over the same area of sky as was used in R12 to measure the power spectrum.

The precision expected for a bispectrum measurement using 800 deg^2 of SPT data should be high enough to provide useful new information. The B12 modeling uncertainty on $A_{\rm tSZ}$ (the amplitude of the tSZ power spectrum) given a perfect measurement of $B_{\rm tSZ}$ (the amplitude of the tSZ bispectrum) is 7% for the default B12 assumptions about gas physics—in which the spread in gas model parameters is constrained by the pressure profile measurements of Arnaud et al. (2010)—and ∼15% when the extreme cases of no feedback and maximal feedback (well beyond the Arnaud et al. 2010 limits) are considered. The predicted scaling between the two quantities is $B_{\rm tSZ} \propto A_{\rm tSZ}^{1.4}$ (B12). This implies that a ∼20% measurement of $B_{\rm tSZ}$ will be limited in its constraint on $A_{\rm tSZ}$ by modeling uncertainty in the pessimistic case, and a ∼10% measurement will be limited by modeling uncertainty in the default case. B12 showed that a survey with the depth and sky coverage of the full 2500-deg^2 SPT-SZ survey should be able to make a ∼6% measurement (or, equivalently, a ∼16σ detection) of the tSZ bispectrum in the 150 GHz data alone. Thus, an ∼11% measurement (∼9σ detection) should be achievable with 800 deg^2 of SPT-SZ 150 GHz data, with the 95 GHz data adding additional detection significance. Depending on the level of modeling uncertainty assumed and the additional uncertainty on $B_{\rm tSZ}$ due to systematics (see Section 4.2.1) and sample variance (see Section 6.3.1), this level of detection has the potential to provide interesting $A_{\rm tSZ}$ constraints.

The data analysis up to the point of generating maps from single observations of each field is identical to that in R12, and we refer the reader to that work and other SPT data analysis papers (e.g., Lueker et al. 2010; Schaffer et al. 2011) for details of the analysis. Briefly, raw, time-ordered detector data from a single observation of a given sky field are relatively calibrated, data selection cuts are applied, high-pass filters are applied to the data to downweight noise from the atmosphere and the readout, and the data are binned into a single-observation map, using simple inverse-variance weighting.

R12 used a cross-spectrum analysis to estimate the power spectrum. This choice, involving the cross-correlation of single-observation maps, was made mainly to avoid noise bias in the power spectrum. In the bispectrum analysis, no noise bias is expected if the instrument/atmospheric noise is Gaussian. For the bispectrum estimation, we therefore use a single coadded map for each field and frequency band, made by taking inverse-variance-weighted averages of all single-observation maps (the total number of single observations ranges from ∼200 to ∼1000). We use simulated observations to characterize the effect of instrument beam and timestream filtering and to estimate bispectrum uncertainties (see Section 3 for details). In this work, we use the single-observation maps and simulation products from the R12 analysis.

The maps used in R12 and in this work are constructed from data taken in the 2008 and 2009 SPT observing seasons. In 2008, only detectors at 220 GHz and 150 GHz produced science-quality data; in 2009, science-quality data was produced in all three frequency bands. The two 2008 fields are ∼100 deg^2 and roughly square on the sky; the three 2009 fields are ∼200 deg^2 and extend roughly twice as far (in real degrees on the sky) in right ascension as they do in declination. To simplify the bispectrum calculation, we split each of the 2009 fields into a left and a right half, each of which is roughly the dimensions of the 2008 fields, leaving us with eight ∼100 deg^2 fields of similar shape. The total sky area analyzed, corrected for any overlap between fields and for regions near bright sources that are interpolated over, is 837 deg^2.

3. BISPECTRUM ESTIMATION METHOD

Previous estimates of non-Gaussianity in CMB data (primordial or otherwise) have generally made use of an estimator characterizing a single amplitude for the non-Gaussian signal. This amplitude parameter is $f_{\rm NL}$ for primordial non-Gaussianity (e.g., Yadav & Wandelt 2008; Komatsu et al. 2011) and $\langle T^3_{\rm tSZ}\rangle$ for non-Gaussianity due to tSZ (Wilson et al. 2012). An alternate analysis method is to calculate the three-point function or bispectrum in full generality, then extract the best-fit amplitude of a given non-Gaussian signal template from the full bispectrum. This more general approach does not require assumptions about the angular scale dependence of the non-Gaussian signal. It also allows the freedom to simultaneously measure different sources of non-Gaussianity, such that signals that are not of interest (for example the bispectrum due to emissive sources in an analysis targeting the primordial CMB bispectrum) can be marginalized over.

Historically a calculation of the full bispectrum has been avoided because it is computationally unfeasible for full-sky data sets (e.g., Yadav & Wandelt 2010). However, over a small patch of sky, one can take advantage of the flat-sky approximation, allowing spherical harmonic transforms to be replaced by fast Fourier transforms (FFTs). In this work, we use the flat-sky approximation and calculate the full, three-dimensional (3D) bispectrum over $\sim$800 deg^2, or roughly 2% of the full sky. Fergusson & Shellard (2009) find that the bispectrum estimated using the flat-sky approximation agrees with the full, curved sky analysis to $\lesssim$1% if all l values are greater than 150. In this work, we only consider multipole values of $l \gtrsim 1000$, so the flat-sky approximation is a very good one for this analysis.

3.1. Defining the Estimator

Following Hu (2000), we define the full-sky (angle-averaged) bispectrum through the relation

$$\langle a_{l_1m_1} a_{l_2m_2} a_{l_3m_3}\rangle = \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{l_1l_2l_3}, \quad (1)$$

where a_{lm} are the coefficients of the spherical harmonic expansion of the full-sky temperature field, and the Wigner 3-j symbol enforces selection rules on the triplets of angular modes. In the flat-sky limit, the equivalent relation is

$$\langle a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)\rangle = (2\pi)^2\,\delta^2(\mathbf{l}_1+\mathbf{l}_2+\mathbf{l}_3)\,B(\mathbf{l}_1,\mathbf{l}_2,\mathbf{l}_3), \quad (2)$$

where $a(\mathbf{l})$ are the coefficients of the Fourier transform of the partial-sky temperature field, defined by

$$\Delta T(\mathbf{x}) = \int \frac{d^2l}{(2\pi)^2} a(\mathbf{l}) e^{i\mathbf{l}\cdot\mathbf{x}}, \quad (3)$$

and the Wigner 3-j symbol has been replaced by a Dirac δ function enforcing that the locations of the three Fourier modes form a triangle in l space. If the signal responsible for the bispectrum has no preferred direction in the sky, we can write this as

$$\langle a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)\rangle = (2\pi)^2\,\delta^2(\mathbf{l}_1+\mathbf{l}_2+\mathbf{l}_3)\,B(l_1,l_2,l_3), \quad (4)$$

where $l_i = |\mathbf{l}_i|$. In this limit, the flat-sky bispectrum $B(l_1, l_2, l_3)$ is equivalent to the full-sky reduced bispectrum $b_{l_1l_2l_3}$, defined through the relation

$$B_{l_1l_2l_3} = \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} b_{l_1l_2l_3} \quad (5)$$

(Komatsu & Spergel 2001).

For a finite-sized map of angular extent $\Delta\theta$, Fourier modes separated by $\Delta l \leqslant 2\pi/\Delta\theta$ are indistinguishable from one another. To avoid significant correlations between mode triplets, the bispectrum of such a map should be calculated in bins of size $\Delta l > 2\pi/\Delta\theta$. Santos et al. (2002) define an estimator of the bispectrum in bins of width Δl:

$$\hat{B}(l_1,l_2,l_3) = \frac{1}{N_{l_1,l_2,l_3;\Delta l}} \times \sum_{l_i-\Delta l/2\leqslant|\mathbf{l}_i|\leqslant l_i+\Delta l/2} \mathrm{Re}\left[a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)\right], \quad (6)$$

subject to the constraint $\mathbf{l}_1+\mathbf{l}_2+\mathbf{l}_3 = 0$, where Re$[x]$ is the real part of the complex number x, and $N_{l_1,l_2,l_3;\Delta l}$ is the number of mode triplets that satisfy the l bounds and the triangle condition. We add the option of assigning different weights to different mode triplets in a bin by redefining the estimator to be

$$\hat{B}(l_1,l_2,l_3) = \frac{1}{\sum W(\mathbf{l}_1,\mathbf{l}_2,\mathbf{l}_3)} \times \sum W(\mathbf{l}_1,\mathbf{l}_2,\mathbf{l}_3)\,\mathrm{Re}\left[a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)\right], \quad (7)$$

where the sum is over the same mode triplets as in Equation (6), and the same triangle condition applies. The weighting scheme used in this analysis is described in Section 3.4. To compare the partial-sky bispectrum estimate to predictions for the full-sky reduced bispectrum, we divide by the map area $\Delta\Omega_{\rm map}$.

We only calculate the auto-bispectrum $\hat{B}(l_1, \nu_i; l_2, \nu_i; l_3, \nu_i)$ for each of our three frequency bands. We are not exploiting the full information in the bispectrum, since there are also seven unique cross-bispectra $\hat{B}(l_1, \nu_i; l_2, \nu_j; l_3, \nu_k)$, where ν_i, ν_j, and ν_k are not all the same. To keep the computation and interpretation as simple as possible for this first result, we postpone investigation of the cross-bispectra to a future publication.

3.2. Treatment of the Instrument Beam and Filter Transfer Function

The maps from which we estimate the bispectrum in this work do not contain unbiased estimates of the true sky temperature at all angular scales, due to the filtering applied to the detector time-ordered data and due to the instrument beam. However, in the limit that the filtering is a purely linear operation that is uniform over the map, we can define a single Fourier-domain function $F(\mathbf{l})$ that describes the combined effects of beam and filtering on the coefficients $a(\mathbf{l})$. We obtain an unbiased estimate of the true $a(\mathbf{l})$ by dividing the raw, biased $\hat{a}(\mathbf{l})$ (estimated by directly Fourier transforming the map) by $F(\mathbf{l})$. We estimate $F(\mathbf{l})$ by taking realistic mock skies (described in detail in Section 3.4.2), convolving them with the measured beam, and running them through the full pipeline up to the coadded map stage. We calculate the two-dimensional (2D) Fourier-domain ratio of output to input maps and use this as our estimate of $F(\mathbf{l})$.

Due to the finite size of the detector array and the sky fields measured, the timestream filtering process cannot truly

be represented by a purely linear map-domain filter that is uniform across the map. However, we expect any errors in the measured transfer function due to this non-linearity or non-uniformity to be very small for two reasons. First, the primary effect of departures from our idealization of the transfer function is to alias power at low spatial frequencies to high spatial frequencies (e.g., Schaffer et al. 2011). This is a potentially significant problem when the signal spectrum is very red (as in measurements of the primary CMB power spectrum), but the expected bispectrum signal in the l range treated here is much closer to flat in l. Second, the input signal we use in the simulations is expected to be a reasonable approximation to the true input signal, which minimizes the impact of nonlinearity. Furthermore, even if the assumed input signal is significantly wrong, the expected errors on the transfer function are small. For example, Lueker et al. (2010) tested the filter transfer function for SPT power spectrum analysis and found that changing the input power from extragalactic sources by a factor of two made $<1\%$ changes in the inferred transfer function.

3.3. Apodization and Compact-source Treatment

We use FFTs to calculate $a(\mathbf{l})$ from our maps, and FFT algorithms assume periodic boundary conditions. To avoid injecting false signals from discontinuities at the edges of the map, we enforce periodic boundary conditions by apodizing each map after embedding it in a larger grid and zero-padding. For a given sky field, we create the apodization mask as follows. We start with a map representing the total inverse-variance weights used in creating the final coadded map for a given field. We smooth the weight map with a Gaussian kernel with FWHM $= 4'$, divide by a fiducial weight value (equal to 80% of the median weights, a value empirically determined to produce well-behaved masks), and set all values above 1.0 equal to 1.0. The resulting apodization mask also downweights the edges of the map, which are noisier than the nearly uniform noise main map region.

The apodization in real space is a convolution in Fourier space, with the convolution kernel being the Fourier transform of the apodization mask. This operation will correlate otherwise uncorrelated Fourier coefficients. However, if the mask has a smooth taper and no features on small angular scales, the Fourier-domain convolution kernel will be compact and of width $\Delta l \simeq 2\pi/\Delta\theta$; i.e., the mode correlation induced will be approximately the same as the correlation that arises just from the finite size of the map. If the l bins used in the bispectrum analysis are sufficiently wide, this correlation can be ignored. We choose a bin size of $\Delta l = 200$; the amplitude of the 2D Fourier transform of the apodization mask for a typical field at $l = 200$ in either dimension is $\lesssim 0.01$ times the $l = 0$ value. We have verified through simulated observations that no detectable correlation between bispectrum values in bins of this size is induced from our apodization masks, and we ignore any effects of the mask—beyond correcting for its effect on $\Delta\Omega_{\rm map}$, the area over which modes are measured—in subsequent analyses.

It is common practice in CMB analyses to also mask compact sources in the maps (e.g., Hinshaw et al. 2003; Lueker et al. 2010; Fowler et al. 2010). Although we are interested in the bispectrum from compact sources such as galaxy clusters, radio sources, and DSFGs, we do want to remove the signal from the very brightest of these sources. Masking the brightest sources reduces sample variance—and, in some cases, uncertainty in modeling their bispectrum—and it allows us to estimate how much of the bispectrum signal is coming from sources that have not been detected and characterized in other analyses of the same data (see Section 4.1.2 for details). However, if we were to multiply the maps by a mask that had holes at source locations, we could no longer ignore the effects of the mask on the measured bispectrum because the mask would now have small-scale features.

To avoid having to calculate the bispectrum equivalent of the Hivon et al. (2002) pseudo-C_l mode-mixing kernel, we instead choose to remove compact-source signals from our maps via harmonic inpainting. We use the procedure described in van Engelen et al. (2012); briefly, a square region around a bright source in the map is interpolated over using the correlation properties measured in the rest of the map to create the interpolates. In all maps, we interpolate over sources detected at 5σ at 150 GHz (using the catalogs of Vieira et al. 2010 and Mocanu et al. 2013). Because the different sky fields used in this analysis were observed to slightly different depths, the 150 GHz flux cut to which this significance is equivalent varies from 5.7 mJy to 6.6 mJy. In some versions of the bispectrum analysis, we also interpolate over galaxy clusters above a given mass from the Reichardt et al. (2013) catalog (see Section 4.1.2 for details). In all cases, the inpainting is done over only $\sim 1\%$ of the total map area. We test for any effects of this inpainting on bispectrum measurements by studying simulated data. We see no effect from inpainting—beyond the obvious effect of eliminating the contribution to the bispectrum from the painted-over sources—at the sensitivity of our tests, which probe down to roughly 1% of the expected secondary/foreground bispectrum signal level.

3.4. Noise and Bispectrum Weighting

The signals we are interested in for this work are non-Gaussian contributions to the sky temperatures recorded in our maps. Any purely Gaussian contributions will, by definition, produce no average bispectrum. However, Gaussian components of the maps will contribute to the variance on the bispectrum measurement. Therefore, in addition to instrumental and atmospheric sources of noise, astrophysical and cosmological sources of Gaussian power (such as the primary CMB and populations of emissive sources) will also contribute noise to this analysis.

When constructing the binned bispectrum, we average together the products of many mode triplets to estimate the bispectrum in each bin. Considering a single mode triplet, the variance on the product of three Fourier coefficients $a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)$ has contributions from the Gaussian components of the map (including the Gaussian part of intrinsically non-Gaussian sources of power such as emissive sources) as well as from the non-Gaussian components. In the limit of very small non-Gaussian signatures in the maps, the Gaussian components dominate this variance.

The three most significant sources of fluctuation power for the maps used in this analysis are noise (instrumental and atmospheric), primary CMB fluctuations, and power from extragalactic sources (mainly DSFGs) below the SPT detection threshold. We measure the SPT noise to be Gaussian at the level necessary for this analysis by calculating the bispectrum of noise-only maps (described in more detail in Section 3.4.1 below) using the bispectrum estimator and obtaining the expected null result. The non-Gaussianity of the primary CMB is constrained through estimates of $f_{\rm NL}$ from *WMAP* data. The signal from extragalactic sources is intrinsically Poisson-distributed, but at flux levels at which we expect many sources per SPT beam,

the distribution of fluxes in a single SPT beam or pixel will approach a Gaussian. Hence the extragalactic source signal will have a non-Gaussian part, which we consider as a potential bispectrum signal, and a Gaussian part, which will contribute to the noise of the bispectrum measurement. The processes used to estimate weights and error bars are described in more detail in Sections 3.4.1 and 3.4.2.

When averaging many mode triplets together, we use weights derived from estimates of the Gaussian variance from noise, primary CMB, and point sources (estimated as described in the next section). Error bars on the binned bispectrum values are also constructed from estimates of the Gaussian variance in the maps. These weights and error bars do not take into account non-Gaussian signal variance. Signal variance for the binned bispectrum can be significant: individual strong sources in the maps produce bispectrum signals that are highly correlated across mode triplets and will vary from field to field across the sky. We take this signal variance into account when interpreting our model fits in a cosmological context, as described in Section 6.3.1.

3.4.1. Weights

In this work, the weights used in the bispectrum estimator described by Equation (7) are constructed from estimates of the Gaussian variance for each mode triplet. A purely Gaussian component with angular power spectrum $C(\mathbf{l})$ will contribute variance to the bispectrum measurement equal to

$$\langle |B(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3)|^2 \rangle = C(\mathbf{l}_1)C(\mathbf{l}_2)C(\mathbf{l}_3) \quad (8)$$

for $\mathbf{l}_1 \neq \mathbf{l}_2 \neq \mathbf{l}_3$, so we use as our bispectrum weights

$$W(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) = \frac{1}{C(\mathbf{l}_1)C(\mathbf{l}_2)C(\mathbf{l}_3)}. \quad (9)$$

Our estimate of the total $C(\mathbf{l})$ contributing to bispectrum variance is the sum of the contributions from primary CMB fluctuations, emissive sources below the SPT detection threshold, and instrumental/atmospheric noise.

The input CMB and point-source spectra are identical to those used in the simulated skies in R12, and we refer to that work for details on the input spectrum. Briefly, the contribution from the primary CMB is a ΛCDM model of the primary CMB power spectrum from Keisler et al. (2011), and the contribution from emissive sources is based on measurements in Shirokoff et al. (2011). The contribution from instrumental and atmospheric noise is estimated using the 2D l-space noise power spectrum associated with the map. The noise power spectrum for a given map is calculated via a jackknife procedure, in which many combinations of the individual-observation maps are created, each one with half the maps multiplied by -1 so that the resulting combination has no astronomical signal. The power spectrum of each of these combinations is computed, and the results are averaged to produce the final estimate of the noise power spectrum $C_{\rm noise}(\mathbf{l})$, which is divided by the square of the beam and transfer function estimate $F(\mathbf{l})$ (see Section 3.2) before being included in the variance calculation.

The expected variance of individual Fourier modes due to instrumental/atmospheric noise and the primary CMB varies across the Fourier plane. The primary CMB variance depends on l, and the noise power spectrum depends on $\mathbf{l}$ (see Schaffer et al. 2011 for details). Our weights take these variations into account.

We address some mode triplets as special cases. For mode triplets in which two of the $\mathbf{l}$ values are the same, the bispectrum variance will be elevated by a factor of three (because $\langle a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)\rangle \rightarrow \langle a^2(\mathbf{l}_1)a(\mathbf{l}_3)\rangle$). Additionally, because the estimator we are using takes the real part of $a(\mathbf{l}_1)a(\mathbf{l}_2)a(\mathbf{l}_3)$, the same noise elevation occurs for mode triplets in which $\mathbf{l}_1 = -\mathbf{l}_2$ or $\mathbf{l}_1 = -\mathbf{l}_3$, etc. Because the fraction of such mode triplets is small (fewer than 0.01% of the total number of triplets over the l range considered here), and they would be significantly downweighted in our weighting scheme, we choose to simply give these triplets zero weight in the estimator.

3.4.2. Binned Bispectrum Error Bars

The total inverse-variance weight in a given l bin is also directly related to the uncertainty on the estimate of the bispectrum in that bin: $\sigma^2(\hat{B}(l_1, l_2, l_3))$ should be equal to $1/W_{\rm tot}$, where $W_{\rm tot}$ is the total weight in that bin

$$W_{\rm tot} = \sum_{l_i - \Delta l/2 \leqslant |\mathbf{l}_i| \leqslant l_i + \Delta l/2} W(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3). \quad (10)$$

In practice, we estimate the bispectrum variance $\sigma^2(\hat{B}(l_1, l_2, l_3))$ from the scatter in the bispectrum measured from 100 simulated observations. As a cross-check, we have compared the binned bispectrum error bars estimated from the weights to those estimated from the simulations. The two are the same up to an overall scaling factor (of order unity) related to the ratio of the area under the apodization mask to the total area of the field, which affects the number of truly independent mode triplets in a bispectrum bin. This decrease in independent modes is due to mode correlation from the mask. This correlation also increases the number of mode triplets with elevated noise (as described in the previous section). However, for masks that are smoothly tapered and cover nearly the full field (such that they are strongly localized in the Fourier domain), the fractional increase in noisy triplets—which were <0.01% of total triplets to begin with (see Section 3.4.1)—is small. Therefore, we ignore this effect in this analysis (i.e., we include these triplets in the estimator and give them the weight they would have in the absence of masking).

We create our final estimate of bispectrum error bars by running 100 sets of mock observations of our eight fields through the bispectrum estimator and calculating the scatter in the measured bispectrum in each l bin across the 100 sets. The input skies are composed of Gaussian realizations of the same sky power spectrum used for the weights described in the previous section, convolved with the measured beam-and-filter transfer function $F(\mathbf{l})$, with a realization of the instrumental/atmospheric noise added. For the simulated noise in a given field, we use one of the signal-free combinations of individual observations used in the noise power spectrum estimation described above. The simulated observations are apodized in the same manner as the real data, so any effects of the apodization are taken into account in the uncertainty estimation. The simulations do not contain correlated signal between fields, so overlap between fields is not taken into account; however, the overlap is ∼2% of the total area, and any error caused by neglecting it is small compared to the statistical precision of our final results.

We have examined the l-bin-to-l-bin covariance over the 100 simulated observations and see no bin-to-bin correlation above the level expected from this number of independent measurements. We treat the noise in each l bin as independent in all subsequent analyses. We do see correlations within an

l bin among the three observing bands, as expected given the contribution to the bispectrum variance from Gaussian sky signal (particularly the primary CMB, which is perfectly correlated among the three bands). We account for these correlations by expressing the bispectrum covariance as a 3×3 matrix in each l bin. We also note that we detect no mean bispectrum in these simulated observations, which include actual instrumental/atmospheric noise, demonstrating that the noise is Gaussian to a very good approximation. The bispectrum covariance matrix used in further analysis is thus

$$C_{ij}(l_1, l_2, l_3) = \langle \hat{B}_{\rm sim}(l_1, l_2, l_3, \nu_i)\, \hat{B}_{\rm sim}(l_1, l_2, l_3, \nu_j) \rangle, \quad (11)$$

where $\hat{B}_{\rm sim}$ is the estimated bispectrum from a single simulation, and the expectation value is over all simulations.

4. BISPECTRUM MODELING AND MODEL FITTING

To interpret any detection of the secondary/foreground bispectrum in an astrophysical or cosmological context, we need a model of the expected signal and a model-fitting procedure. In this section, we describe the signal models we adopt and the procedure we use for fitting the multi-band data to these models. We also describe how we account for instrument-related systematic effects such as uncertainties in beams, spectral bandpasses, and calibration.

4.1. Signal Models

We include three types of non-Gaussian signal in our modeling: tSZ from galaxy clusters, the spatially uncorrelated signal from extragalactic sources (hereafter "point sources," since the vast majority of such sources will appear point-like at the $\sim 1'$ resolution of the SPT), and the expected clustered emission from one source population (DSFGs). We describe our modeling choices for each of these in turn, but first we note that we do not include other potential sources of millimeter-wave signal—particularly the kSZ effect, clustered radio sources, and galactic foregrounds—in the bispectrum model.

We do not include the kSZ primarily because none of the predicted kSZ-generating mechanisms—including the peculiar velocity of free electrons in galaxies (Ostriker & Vishniac 1986) or galaxy clusters (Sunyaev & Zel'dovich 1980), and patchy reionization (Knox et al. 1998; Gruzinov & Hu 1998)—is expected to impart a net skewness on the CMB temperature distribution. This is because the velocities of ionized gas in the universe should be random with respect to the observer, meaning that the induced kSZ signal should be symmetric around the mean CMB temperature. We expect the signal from the clustered radio background to be much smaller than the clustered DSFG signal (see Section 4.1.3 for details), and we do not include such a term in our signal model. We do not include any expected signal from our own galaxy, both because the sky fields used here are at high galactic latitude, and because such signals are expected to fall steeply with l (e.g., Finkbeiner et al. 1999) and thus be negligible at the angular scales or l values of interest to this work.

4.1.1. Spatially Uncorrelated ("Poisson") Point-source Contribution

We introduce the model for spatially uncorrelated point sources first to illustrate the basic properties of the bispectrum arising from any population of discrete, spatially uncorrelated sources with a given angular profile. Following, e.g., Hall et al. (2010), we will use "Poisson" as shorthand for the component of the point-source contribution to the power spectrum or bispectrum that arises from spatially uncorrelated sources.

For a CMB map made at observing frequency ν with pixels of size $\Delta\Omega_p$ (where, for this toy example, the pixel size is much larger than the beam size), containing only a point source of flux S, we can write the signal in the map as

$$T_{\rm source}(\mathbf{x}) = \begin{cases} T_{\rm peak}, & \text{if } \mathbf{x} \in \text{source pixel} \\ 0, & \text{otherwise} \end{cases}. \quad (12)$$

We know that the total flux in the map must equal S, which means

$$T_{\rm peak}(\nu) = g(x_\nu) \times \frac{S}{\Delta\Omega_p}, \quad (13)$$

where $g(x_\nu)$ is the conversion factor between CMB fluctuation temperature and flux per solid angle (in units of Jy sr^{-1}) at observing frequency ν:

$$g(x_\nu) = 10^{-26} \times \left[\frac{2k_B}{c^2} \left(\frac{k_B T_{\rm CMB}}{h} \right)^2 \frac{x_\nu^4 e^{x_\nu}}{(e^{x_\nu} - 1)^2} \right]^{-1}, \quad (14)$$

and $x_\nu = h\nu/(k_B T_{\rm CMB})$. For angular frequencies well below the cutoff of the pixel window function ($l \ll 2\pi/\sqrt{\Delta\Omega_p}$), the Fourier transform of this map is

$$\begin{aligned} a_{\rm source}(\mathbf{l}) &= \int d^2x\; T_{\rm source}(\mathbf{x})\, e^{-i\mathbf{l}\cdot\mathbf{x}} \\ &\simeq \Delta\Omega_p\; T_{\rm peak}\; e^{-i\mathbf{l}\cdot\mathbf{x}_{\rm source}} \\ &= g(x_\nu)\; S\; e^{-i\mathbf{l}\cdot\mathbf{x}_{\rm source}}. \end{aligned} \quad (15)$$

The estimated bispectrum due to this single source of flux S is then

$$\begin{aligned} \hat{B}(l_1, l_2, l_3, \nu) &= \frac{1}{N_l} \sum_l \mathrm{Re}\left[a(\mathbf{l}_1) a(\mathbf{l}_2) a(-(\mathbf{l}_1 + \mathbf{l}_2))\right] \\ &= \frac{g^3(x_\nu)\; S^3}{N_l} \sum_l \mathrm{Re}\left[e^{-i\mathbf{l}_1\cdot\mathbf{x}_{\rm source}} e^{-i\mathbf{l}_2\cdot\mathbf{x}_{\rm source}} \right. \\ &\quad \left. \times\, e^{i(\mathbf{l}_1+\mathbf{l}_2)\cdot\mathbf{x}_{\rm source}}\right] \\ &= g^3(x_\nu)\; S^3, \end{aligned} \quad (16)$$

where we have used the triangle condition $\mathbf{l}_1 + \mathbf{l}_2 + \mathbf{l}_3 = 0$ to redefine $\mathbf{l}_3$. When there are two or more sources in the map, Equation (16) becomes

$$\begin{aligned} \hat{B}(l_1, l_2, l_3, \nu) = g^3(x_\nu) \left(S_1^3 + S_2^3 + \cdots + S_N^3 \right. \\ \left. + \text{cross terms}\right), \end{aligned} \quad (17)$$

where the cross terms are of the form $S_1^2 S_2 e^{-i(\mathbf{l}_1+\mathbf{l}_2)\cdot(\mathbf{x}_1-\mathbf{x}_2)}$. If the sources are spatially uncorrelated, the phase of the cross terms is random, and these terms will on average be zero, leaving

$$\hat{B}(l_1, l_2, l_3, \nu) = g^3(x_\nu) \left(S_1^3 + S_2^3 + \cdots + S_N^3\right) \quad (18)$$

as the only nonzero average bispectrum contribution.

For a population of sources with number density per unit solid angle per unit flux $dN/dS/d\Omega$, Equation (18) is easily generalized to

$$\hat{B}(l_1, l_2, l_3, \nu) = g^3(x_\nu)\, \Delta\Omega_{\rm map} \int_0^\infty S^3 \frac{dN}{dS d\Omega} dS, \quad (19)$$

or, if sources have been cleaned down to some threshold $S_{\rm max}$,

$$\hat{B}(l_1, l_2, l_3, \nu) = g^3(x_\nu)\, \Delta\Omega_{\rm map} \int_0^{S_{\rm max}} S^3 \frac{dN}{dS d\Omega} dS, \quad (20)$$

which, after we apply the $\Delta\Omega_{\rm map}$ correction, is identical to the familiar result of, e.g., Komatsu & Spergel (2001).

We have chosen to only use bispectrum shape information to fit for the Poisson contribution to our multi-band bispectrum (i.e., we fit for a contribution that is flat in l). We include a single free parameter in the fit for the amplitude of the Poisson source bispectrum in each observing band.

4.1.2. Thermal SZ Model

We use the model described in B12 for the bispectrum due to the tSZ effect in galaxy clusters. We describe the model briefly here and refer the reader to B12 for details. The tSZ bispectrum is calculated assuming the signal arises from spatially uncorrelated galaxy clusters and so is conceptually identical to the result from the previous section, with the modification that the intrinsic angular profile of the clusters modifies the bispectrum shape. For a family of astrophysical sources with angular profile $F(\mathbf{x})$ or Fourier-domain profile $F(\mathbf{l})$, $a_{\rm source}(\mathbf{l}) \rightarrow F(\mathbf{l})\, a_{\rm source}(\mathbf{l})$, and the bispectrum $B(l_1, l_2, l_3) \rightarrow F(l_1)F(l_2)F(l_3)B(l_1, l_2, l_3)$.

The tSZ bispectrum at multipole numbers l_1, l_2, and l_3 and observing frequency ν is calculated as the integral over cosmological volume of the product of the Fourier-domain cluster pressure profile at the three l values, weighted by the halo mass function:

$$B_{\rm tSZ}(l_1, l_2, l_3, \nu) = f^3(x_\nu) \int dz \frac{dV}{dz} \int d\ln M \frac{dn(M, z)}{d\ln M} \times y(M, z, l_1) y(M, z, l_2) y(M, z, l_3), \quad (21)$$

where $f(x_\nu)$ is the dimensionless function specifying the dependence of the tSZ on observing frequency (Sunyaev & Zel'dovich 1980). We do not include relativistic corrections to $f(x_\nu)$ (see discussion below). The Fourier-domain pressure profile $y(M, z, l)$ is calculated from the analytic model of Shaw et al. (2010), using their fiducial values of the intracluster medium (ICM) parameters. The halo mass function $dn(M, z)/d\ln M$ is from Tinker et al. (2008). A ΛCDM cosmology is assumed in calculating the halo mass function, with fiducial parameters as in B12, namely $\Omega_b = 0.045$, $\Omega_m = 0.27$, $h = 0.71$, $n_s = 0.97$, and $\sigma_8 = 0.8$.

The model is calculated at the center of each l bin in which the bispectrum is estimated from the data. The signal is sufficiently flat in l that this is within 2% of the value that would be obtained by calculating the model at higher resolution in l and averaging over the bin. The tSZ frequency factor $f(x_\nu)$ is evaluated for each observing band at the effective center frequency of the band assuming a non-relativistic tSZ spectrum. R12 calculated these frequencies to be 97.6, 152.9, and 218.1 GHz. (This value is an average over the 2008 and 2009 observing seasons for the 150 and 220 GHz bands, see R12 for details.)

There is some uncertainty in how well mass function fits to simulation output capture the high-mass end, with potential 5%–10% uncertainties at halo masses above $\sim 10^{15}\, M_\odot$ (Tinker et al. 2008; Bhattacharya et al. 2011). For this reason, we also calculate a version of the tSZ model with the mass function truncated above $M_{200}(\rho_{\rm crit}) = 8 \times 10^{14}\, M_\odot\, h^{-1}$, where $M_{200}(\rho_{\rm crit})$ is the mass enclosed inside $R_{200}(\rho_{\rm crit})$, defined as the radius within which the average density is 200 times the critical density. To compare to this prediction, we construct a bispectrum estimate in all three SPT bands with clusters above this same mass removed from the maps (using the same inpainting procedure used for the point sources, see Section 3.3). The cluster masses are taken from the Reichardt et al. (2013) catalog and converted from M_{500} to M_{200} assuming a Navarro et al. (1996) profile and the Duffy et al. (2008) mass-concentration relation. A total of four clusters above this threshold are masked in the full ∼800 deg^2 data set.

This level of cluster masking also reduces potential systematic errors caused by ignoring relativistic corrections to the predicted tSZ bispectrum amplitude. The most massive, hottest clusters have gas temperatures of $\gtrsim$10 keV (e.g., Allen et al. 2008). At these temperatures, the relativistic correction to the tSZ temperature decrement is ∼6% at 150 GHz (Nozawa et al. 2000). Limiting the cluster sample to $M_{200}(\rho_{\rm crit}) < 8 \times 10^{14}\, M_\odot\, h^{-1}$ is roughly equivalent to a temperature limit of $T < 5$ keV (Stanek et al. 2010). At these temperatures, the maximum error in $f(x_\nu)$ from ignoring relativistic corrections is ∼3%.

We also construct tSZ models and bispectrum estimates with the mass function truncated and clusters masked above $M_{200}(\rho_{\rm crit}) = 3 \times 10^{14}\, M_\odot\, h^{-1}$. This mass threshold closely approximates the selection of clusters used to constrain cosmology in Reichardt et al. (2013), namely signal-to-noise ratio greater than five and $z \geqslant 0.3$. This allows us to estimate the amount of information we are extracting from the tSZ bispectrum above and beyond what has already been extracted using cluster counts. A total of 117 clusters above this threshold are masked in the full ∼800 deg^2 data set. For comparison, the total number of clusters used in the Reichardt et al. (2013) cosmological results was 100.

Masking clusters in the data will lead to a smaller absolute amplitude of the tSZ bispectrum but will also lead to smaller sample variance, since the sample variance is dominated by the presence or absence in a map of the rarest, most massive clusters. On the other hand, any systematic uncertainty in the method used to estimate cluster masses will lead to an uncertainty in the true mass threshold used for masking, resulting in a systematic uncertainty when comparing the masked data to a tSZ bispectrum model (in which the mass threshold for masking is known perfectly). The different scalings with cluster mass of these various contributions to the tSZ bispectrum error budget imply that there may be some optimal mass cut that reduces the combined statistical-plus-systematic-plus-sample-variance uncertainty on the tSZ bispectrum, similar to the results in Shaw et al. (2009) for the tSZ power spectrum. We investigate this further in Section 6.3.1.

4.1.3. Clustered CIB Model

Not only will emissive sources contribute to the Poisson bispectrum, they can also be spatially clustered, possibly leading to a detectable bispectrum signal with a different shape. Because this signal arises from a spatial modulation of the mean intensity, and because the CIB is much brighter than the radio background at SPT observing frequencies (e.g., Hauser & Dwek 2001), we expect the clustering signal from DSFGs to be much stronger than that from radio sources, as has been found in power spectrum measurements (e.g., Hall et al. 2010; Holder 2014). We do not include the clustered radio background in our modeling and concentrate on the potentially measurable signal from the clustered CIB.

As pointed out in Lacasa (2012), a single population of sources with clustering properties described by a single

correlation function or angular power spectrum will have a bispectrum equal to

$$B_{\rm tot}(l_1, l_2, l_3) = \alpha\sqrt{C_{\rm tot}(l_1)C_{\rm tot}(l_2)C_{\rm tot}(l_3)}, \quad (22)$$

where α is a constant, and $C_{\rm tot}$ is the total Poisson-plus-clustering angular power spectrum: $C_{\rm tot} = C_{\rm clust} + C_{\rm poiss}$. Lacasa (2012) further showed that this formulation provides an accurate characterization of the CIB bispectrum in the simulated sky maps of Sehgal et al. (2010).

In this formulation, for l triplet bins in which the clustering signal dominates over the Poisson contribution, we can write the clustering signal as

$$B_{\rm clust}(l_1, l_2, l_3) \propto \sqrt{C_{\rm clust}(l_1)C_{\rm clust}(l_2)C_{\rm clust}(l_3)}. \quad (23)$$

We use this as the l-space template for the clustered CIB bispectrum in our model fits, with a simple power-law model for the clustered power spectrum $C_{\rm clust}(l) \propto l^{-n}$. Both Addison et al. (2012) and R12 have found this to be a good description of the clustered CIB power spectrum over a large range of angular scales, including the entire range considered in this work. We choose $n = 1.2$, which is consistent with the best-fit values in Addison et al. (2012) and R12.

The spectral behavior of the clustered CIB at millimeter wavelengths is fairly well constrained from power spectrum measurements, and we use existing results to inform our model fitting. We assume a single spectral index $\alpha_{\rm c}$ over our three observing bands and model the clustered CIB bispectrum at observing frequency ν as

$$\begin{aligned} B_{\rm clust}(l_1, l_2, l_3, \nu) &= B_{\rm clust}(l_1, l_2, l_3, \nu_0)\left(\frac{\nu}{\nu_0}\right)^{3\alpha_{\rm c}} \\ &= B^{2000}_{\rm clust}(\nu_0)\left(\frac{\nu}{\nu_0}\right)^{3\alpha_{\rm c}} \\ &\quad \times \left(\frac{l_1}{2000}\frac{l_2}{2000}\frac{l_3}{2000}\right)^{-n/2}, \end{aligned} \quad (24)$$

where $B^{2000}_{\rm clust}(\nu_0) = B_{\rm clust}(l_1 = 2000, l_2 = 2000, l_3 = 2000, \nu_0)$. We use $\nu_0 = 220$ GHz in our CIB model.

Using the results of R12, we adopt a nominal value and 1σ uncertainty for the clustered CIB spectral index of $\alpha_{\rm c} = 3.72 \pm 0.12$ (see Section 4.2.1 for how this uncertainty is included in the fit). In Equation (24), ν for each observing band is the effective center frequency of the band assuming a $\nu^{\alpha_{\rm c}}$ spectrum. R12 calculated these frequencies to be 97.9, 153.8, and 219.6 GHz for $\alpha_{\rm c} = 3.5$; the difference in effective frequencies for $\alpha_{\rm c} = 3.5$ and $\alpha_{\rm c} = 3.7$ is negligible.

We have chosen not to explore more complicated CIB modeling, involving (for instance) spatial correlations between the sources of the tSZ and CIB bispectra, different CIB bispectrum shapes, and spectral behavior beyond a single spectral index. As will be shown in Section 6.2, the simple model adopted here is adequate to describe the data, and extensions to this model would not be strongly constrained using the data in this work.

For the particular case of tSZ–CIB correlation, we note that this effect should be a far less significant bias to the measurement of the tSZ bispectrum than it is to the tSZ power spectrum. Galaxies in the high-mass, low-redshift clusters that make up the bulk of the tSZ bispectrum signal are measured to have significantly less star formation per unit mass than galaxies in lower-mass clusters or low-redshift field galaxies (e.g., Hashimoto et al. 1998). Furthermore, the star-forming fraction is also seen to increase with redshift (e.g., Butcher & Oemler 1984). This evidence all indicates that, even if tSZ–CIB correlation has a significant effect on the tSZ power spectrum—which is sourced by lower-mass, higher-redshift clusters than the bispectrum—the tSZ bispectrum is unlikely to be significantly affected. When R12 allow tSZ–CIB correlation as a free parameter, the best-fit tSZ power spectrum amplitude shifts by $<20\%$, so we assume that the effect of tSZ–CIB correlation on our measurement of the tSZ bispectrum will be $\ll 20\%$ and hence subdominant to other uncertainties.

4.2. Fitting Procedure

We use a simple linear least-squares procedure (e.g., Press et al. 1986) to fit the measured bispectrum with the three-component model described above. Least-squares fitting results in the maximum-likelihood estimate of model parameters only if the measurement uncertainties are Gaussian-distributed. While the distribution of the individual l-space mode triplets is highly non-Gaussian, each $\Delta l = 200$ bin contains $>10^4$ of these triplets, so we expect the distribution of binned bispectrum uncertainties to be very nearly Gaussian. We confirm this through simulations.

We fit all three bands' bispectrum data simultaneously. The data vector has $3 \times N_{\rm bin}$ elements, where $N_{\rm bin} = (l_{\rm max}/\Delta l)^3$, $l_{\rm max}$ is the maximum angular frequency used in the fit and Δl is the size of the bins in l space, in this case 200. None of the signal models described in the previous section have features on the scale of $\Delta l = 200$, so this resolution should be adequate to characterize the measured bispectrum. The maximum l used in this analysis is 11,000, which was chosen by investigating the factor by which the bispectrum variance is inflated by deconvolving the beam and transfer function $F(\mathbf{l})$ from the maps. The raw SPT map noise at high l has a nearly white spectrum. After deconvolving $F(\mathbf{l})$, the noise power spectrum of the maps will be proportional to $1/F^2(\mathbf{l})$ at high l. The bispectrum variance from this map will thus be proportional to $1/F^6(\mathbf{l})$ at high l. This factor $F^6(\mathbf{l})$ is >500 for all SPT bands at $l > 11{,}000$.

We write the data vector as d_μ where the index μ is the product of an l-bin index a and an observing-frequency index i (such that μ takes on a unique value for each bin $l_\alpha, l_\beta, l_\gamma$ and observing frequency ν_i):

$$\begin{aligned} d_\mu &= d_{[ia]} = d_{[ia\beta\gamma]} = \hat{B}(l_\alpha, l_\beta, l_\gamma, \nu_i), \\ a &= \alpha N^2_{\rm bin} + \beta N_{\rm bin} + \gamma, \\ \mu &= 3 \times a + i. \end{aligned} \quad (25)$$

The weight matrix, which is the inverse of the bin–bin–band–band bispectrum covariance matrix, is assumed to be block-diagonal in this analysis—i.e., we assume no bispectrum covariance between l bins due to noise or Gaussian sky components, but we do include the band–band covariance of the Gaussian sky terms (see Section 3.4). Under this assumption, each sub-matrix characterizing the covariance between bands for a given bispectrum bin is an independent 3×3 matrix given by the inverse of the band–band covariance matrix for that bin. This covariance matrix is estimated from simulations, as described in Section 3.4. Thus, we can write the weight and covariance matrices as

$$\begin{aligned} W_{\mu\nu} &= C^{-1}_{\mu\nu} \\ C_{\mu\nu} &= C_{[ia][jb]} = C_{[ia][ja]}\delta_{ab}, \end{aligned} \quad (26)$$

where $C_{[ia][ja]} = C_{[i\alpha\beta\gamma][j\alpha\beta\gamma]} = C_{ij}(l_\alpha, l_\beta, l_\gamma)$, as defined in Equation (11), and, again, the indices i and j run over observing bands and the remaining indices over l-space bins.

The model or design matrix A is composed of five $3 \times N_{\rm bin}$-element vectors, each representing the unnormalized signal shape for one of the signal components in all observing bands. The tSZ and clustered CIB vectors have non-zero values in all observing bands (although the model amplitude for the tSZ in the 220 GHz band is very small, since that band is very near the tSZ null), while the three vectors representing the Poisson point-source power in each band (assumed to be independent in this fit) are non-zero only in the $N_{\rm bin}$ elements corresponding to that band.

The five free parameters of the model λ are the amplitudes for each model component: tSZ, clustered CIB, and the Poisson point-source component in each of three bands. The best-fit values of these parameters are estimated from the data as

$$\overline{\lambda_\psi} = \left(A^T_{\psi\mu} W_{\mu\nu} A_{\nu\omega}\right)^{-1} A^T_{\omega\pi} W_{\pi\rho} d_\rho, \quad (27)$$

where sums over repeated indices are assumed. This estimate of parameters has a covariance matrix equal to

$$C^{\rm param}_{\psi\omega} = \langle \delta\lambda_\psi \, \delta\lambda_\omega \rangle = \left(A^T_{\psi\mu} W_{\mu\nu} A_{\nu\omega}\right)^{-1}. \quad (28)$$

4.2.1. Incorporating Systematic Uncertainties

The disadvantage of a simple linear least-squares fit (in comparison to a more general parameter-space search such as a Markov Chain Monte Carlo) is that there is no way to trivially incorporate systematic uncertainties in such quantities as the instrument beam measurement or the spectral index of clustered CIB fluctuations without introducing strong covariance among all l bins and significantly complicating the inversion of the covariance matrix. To retain the advantages of the linear fit (speed, simplicity, and robust parameter covariance estimation), we account for such systematics by running the linear fit many times, each time using a different realization of each systematic effect. We then calculate a systematic parameter covariance matrix $C^{\rm syst}_{\psi\omega}$ by directly computing the outer product $\delta\lambda_\psi \, \delta\lambda_\omega$ in each realization and averaging over all realizations.

We account for four independent sources of systematic uncertainty: (1) instrument spectral bandpasses, (2) the spectral index of CIB fluctuations $\alpha_{\rm c}$, (3) instrument calibration, and (4) instrument beams. Based on measurements described in Schaffer et al. (2011) and similar measurements in 2009, the band centers for SPT are estimated to be accurate to 0.3 GHz. The major source of this uncertainty is the frequency calibration of the Fourier transform spectrometer used to measure the bandpasses, implying that the uncertainty should be highly correlated between the three bands. For each systematic realization, we draw a bandpass error from a Gaussian of width $\sigma_{\rm band} = 0.3$ GHz and calculate the signal models using band centers shifted by this error. To account for uncertainty in the spectral behavior of the CIB, in each systematic realization, we draw a value for $\alpha_{\rm c}$ from the R12 distribution $\alpha_{\rm c} = 3.72 \pm 0.12$ and use that value in calculating the clustered CIB model.

R12 estimated the calibration uncertainty in the three bands to be 0.035, 0.032, and 0.048 in power, or 0.018, 0.016, and 0.024 in temperature. These uncertainties are highly correlated, because a primary source of uncertainty in each band's calibration is the noise in the *WMAP* power spectrum in the range $l \in [650, 1000]$. We approximate the calibration covariance matrix by assigning the fractional uncertainty at 150 GHz to all bands as a fully correlated component and augmenting that with uncorrelated components at 95 and 220 GHz to make the on-diagonal elements equal to the square of the measured uncertainties in each band. For each systematic realization we create a three-element vector $\sigma_{\rm cal}(\nu)$ with the appropriate covariance and multiply the elements of the data vector d corresponding to band ν by $[1 + \sigma_{\rm cal}(\nu)]^3$. The mean and 1σ width of the systematic distributions for bandpass, CIB spectral index, and calibration errors are summarized in Table 1.

Table 1
Systematic Error Accounting

Parameter	Nominal Value	1σ Uncertainty
tSZ band center, 95 GHz	97.6 GHz	0.3 GHz
tSZ band center, 150 GHz	152.8 GHz	0.3 GHz
tSZ band center, 220 GHz	219.1 GHz	0.3 GHz
CIB band center, 95 GHz	97.9 GHz	0.3 GHz
CIB band center, 150 GHz	153.8 GHz	0.3 GHz
CIB band center, 220 GHz	219.6 GHz	0.3 GHz
CIB spectral index	3.72	0.12
Calibration, 95 GHz	1.00	0.018
Calibration, 150 GHz	1.00	0.016
Calibration, 220 GHz	1.00	0.024

Notes. Distributions from which the systematic error realizations described in Sections 4.2.1 and 5.2 are applied. The band center and calibration uncertainties are highly correlated between bands. Uncertainty in the instrument beam in each frequency band is also taken into account using realizations of "error beams" as described in the text.

Uncertainties in the measurement of the instrument beam are incorporated by creating realizations of the beams using the full beam covariance matrix described in Keisler et al. (2011). For each systematic realization, a beam realization is created for each observing band and observing season, including the correlations in the uncertainties between bands and seasons. The bispectrum estimate from each 100 deg^2 field and each band is multiplied by the cube of the ratio of the appropriate beam realization (for the year the field was observed) to the nominal beam. The data from all fields are then combined using the nominal weights, and this combined beam-error-multiplied bispectrum is used to construct the data vector d.

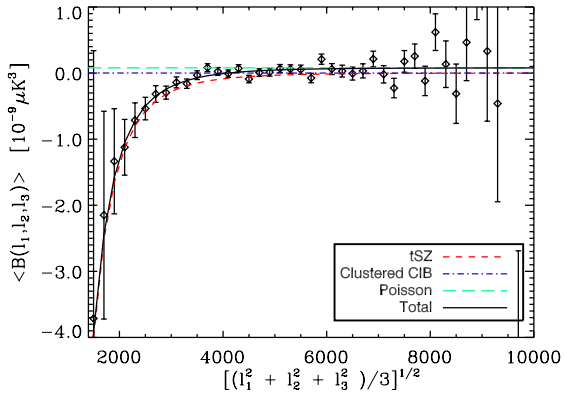
5. RESULTS

5.1. Measured Bispectra and Single-band Detection Significance

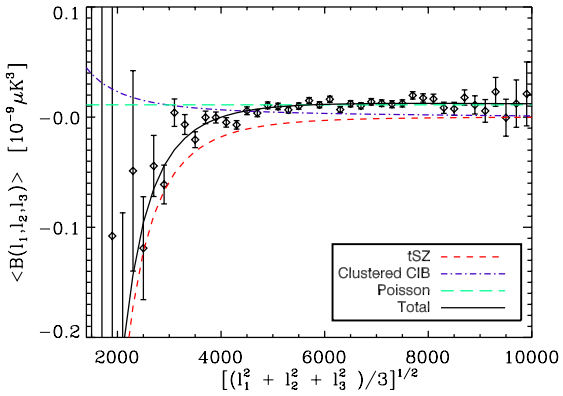
The bispectrum in each frequency band (with no cluster masking), as estimated using the analysis procedures detailed in Section 3 and the bispectrum estimator discussed in Section 3.1, is plotted in Figure 1. Values of the bispectrum and inverse-variance weight in each band and each l_1, l_2, l_3 bin are available for download from the SPT Web site.[24]

We note that displaying this inherently 3D data product in one dimension (1D) requires the data to be contracted along two axes. There is no "industry standard" for displaying bispectra, particularly real measurements with noise. B12 used the "skewness spectrum" $\Lambda(l) = \sqrt{\sum_{l_1,l_2} b^2(l, l_1, l_2)}$; however, this quantity will have a positive expectation value for a bispectrum estimated from data with zero non-Gaussianity but finite noise and Gaussian sky power. We choose to define an

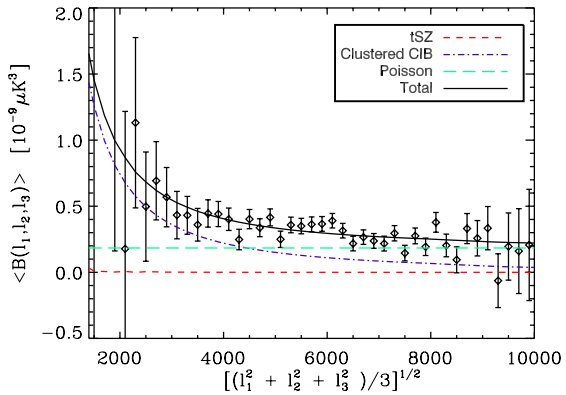
[24] http://pole.uchicago.edu/public/data/crawford13



(a) 95 GHz 1d bispectrum, with best-fit model overplotted



(b) 150 GHz 1d bispectrum, with best-fit model overplotted



(c) 220 GHz 1d bispectrum, with best-fit model overplotted

Figure 1. Measured bispectrum (with no clusters masked) in each of the three SPT frequency bands. Bispectra have been collapsed from three dimensions to one dimension as described in the text. The solid black line shows the best-fit model estimated from the full 3D bispectrum (collapsed to 1D using the same procedure and weighting as used for the data). The three individual components of the best-fit model are also plotted: tSZ (short-dashed red line), clustered CIB (dot–dashed purple line), and the Poisson point-source component (long-dashed green line). See Section 4 for more details on the model. The bispectrum error bars shown are statistical only. The data and best-fit models shown are for the nominal values of the systematic parameters and with no cluster masking (see text for details).

(A color version of this figure is available in the online journal.)

l-space radius

$$l_{\rm rad} = \sqrt{\frac{l_1^2 + l_2^2 + l_3^2}{3}} \tag{29}$$

and to plot

$$\hat{B}(l_{\rm rad}) = \frac{\sum_{l_1,l_2,l_3 \in l_{\rm rad}} W(l_1,l_2,l_3)\hat{B}(l_1,l_2,l_3)}{\sum_{l_1,l_2,l_3 \in l_{\rm rad}} W(l_1,l_2,l_3)}, \tag{30}$$

where $W(l_1, l_2, l_3)$ are the bispectrum weights in an l bin defined in Section 3.4. The error bar on this 1D quantity is

$$\sigma(\hat{B}(l_{\rm rad})) = \sqrt{\frac{1}{\sum_{l_1,l_2,l_3 \in l_{\rm rad}} W(l_1,l_2,l_3)}}. \tag{31}$$

We emphasize that this contraction to 1D is only for display purposes; all model fitting and χ^2 estimation is performed in the full 3D l space. However, when we calculate $B_{\rm tSZ}$ from the B12 model to study the cosmological scaling and modeling uncertainties, we do use the value of this 1D quantity at $l_{\rm rad} = 3000$ as a convenient proxy, rather than performing the full 3D fit—see Section 6.3.3 for details.

Three features of the bispectrum data are immediately clear from Figure 1: (1) the data are highly inconsistent with zero bispectrum in all bands; (2) all bands show evidence of two signal components, namely a component that is larger at low $l_{\rm rad}$ than high $l_{\rm rad}$ and is roughly consistent with a power law in $l_{\rm rad}$, and a flat-in-$l_{\rm rad}$ component consistent with a Poisson point-source component (note that a signal that is flat in l will also be flat in $l_{\rm rad}$); and (3) the power-law component is negative at 95 and 150 GHz but positive at 220 GHz, as would be expected from a bispectrum dominated by tSZ at 95 and 150 GHz and by clustered CIB at 220 GHz.

The results of fitting this data using the multi-frequency model from Section 4 are discussed below. However, to make a reasonably model-independent statement about the preference for these two components in the data, we first fit each band's data individually to a toy model that includes a flat component and a simple power law in l, with a power-law index chosen to match the observed signal in all bands. This turns out to be roughly $B(l_1, l_2, l_3) \propto (l_1 l_2 l_3)^{2/3}$ or $B \propto l^2$ in the equilateral configuration ($l_1 = l_2 = l_3$). Both components are detected strongly in all three bands, with the significance of the power-law component ranging from 5σ at 220 GHz to 9σ at 150 GHz.

Table 2
Multi-band Fit Results, No Cluster Masking

Template	Best-fit Value	Noise Error (1σ)	Systematic Error (1σ)	Quadrature Sum	Detection Significance	Constraint Significance
tSZ, relative to analytical prediction	0.53	0.04	0.03	0.05	13	10
Clustered CIB [10^{-9} μK^3 at $l = 2000$ and $\nu = 220$ GHz]	0.68	0.13	0.06	0.14	5.1	4.7
Poisson, 95 GHz, [10^{-10} μK^3]	0.79	0.16	0.05	0.16	5.0	4.8
Poisson, 150 GHz, [10^{-11} μK^3]	1.12	0.09	0.09	0.13	12	8.7
Poisson, 220 GHz, [10^{-10} μK^3]	1.84	0.26	0.23	0.35	7.0	5.3

Notes. Parameter best-fit values and 1σ statistical and systematic uncertainties. "Detection significance" refers to the best-fit parameter value divided by statistical uncertainty only; "constraint significance" refers to the best-fit parameter value divided by the quadrature sum of statistical and systematic uncertainty. Sample variance is not included in the constraint significance; see Section 6.3 and Table 8 for tSZ results including sample variance.

Table 3
Multi-band Fit Results, $M_{200}(\rho_{\rm crit}) > 8 \times 10^{14}\ M_\odot\ h^{-1}$ Clusters Masked

Template	Best-fit Value	Noise Error (1σ)	Systematic Error (1σ)	Quadrature Sum	Detection Significance	Constraint Significance
tSZ, relative to analytical prediction	0.59	0.05	0.04	0.07	11	9.0
Clustered CIB [10^{-9} μK^3 at $l = 2000$ and $\nu = 220$ GHz]	0.74	0.13	0.07	0.15	5.6	5.0
Poisson, 95 GHz, [10^{-10} μK^3]	0.88	0.16	0.05	0.17	5.6	5.3
Poisson, 150 GHz, [10^{-11} μK^3]	1.25	0.10	0.10	0.14	13	9.1
Poisson, 220 GHz, [10^{-10} μK^3]	1.73	0.26	0.21	0.34	6.6	5.2

Note. See Table 2 for notes.

Table 4
Multi-band Fit Results, $M_{200}(\rho_{\rm crit}) > 3 \times 10^{14}\ M_\odot\ h^{-1}$ Clusters Masked

Template	Best-fit Value	Noise Error (1σ)	Systematic Error (1σ)	Quadrature Sum	Detection Significance	Constraint Significance
tSZ, relative to analytical prediction	0.68	0.17	0.06	0.18	4.0	3.8
Clustered CIB [10^{-9} μK^3 at $l = 2000$ and $\nu = 220$ GHz]	0.75	0.13	0.07	0.15	5.6	4.9
Poisson, 95 GHz, [10^{-10} μK^3]	0.93	0.16	0.06	0.17	5.7	5.4
Poisson, 150 GHz, [10^{-11} μK^3]	1.27	0.10	0.10	0.14	13	9.0
Poisson, 220 GHz, [10^{-10} μK^3]	1.67	0.27	0.20	0.33	6.3	5.0

Note. See Table 2 for notes.

To assess whether the data still prefer a power-law bispectrum component with the most significantly detected clusters masked, we estimate the bispectrum in each band while masking all clusters with $M_{200}(\rho_{\rm crit}) > 3 \times 10^{14}\ M_\odot\ h^{-1}$, which is very close to a cut at signal-to-noise of five in the Reichardt et al. (2013) catalog. With this level of masking, the detection significance of the power-law component at 95 and 150 GHz data is much reduced but still 1σ–2σ in each band.

Perhaps most intriguing is the detection of a power-law component in the 220 GHz data, which is near the tSZ null and should not be measuring a tSZ bispectrum. We interpret this signal as the bispectrum of the clustered CIB, and we discuss the implications of this signal in Section 6.2.

5.2. Results of Model Fits

Having established that the bispectrum data in each band contain significant detections of a power-law component and a flat-in-l component, we move on to fitting these data to the model described in Section 4.1, using the fitting procedure described in Section 4.2. As described in Section 4.2.1, the linear least-squares fit is repeated many times with different realizations of systematic uncertainties, drawn from distributions summarized in Table 1, or, in the case of the instrument beam uncertainties, using beam realizations described in Section 4.2.1. To minimize uncertainty in interpreting the tSZ result due to uncertainties in the assumed halo mass function, we repeat the fit using data in which all clusters above $M_{200}(\rho_{\rm crit}) = 8 \times 10^{14}\ M_\odot\ h^{-1}$ masked and a tSZ model template with the mass function truncated at that value. To determine how much of the tSZ bispectrum is coming from clusters not already used for cosmological constraints from cluster count analyses, we repeat the fit using data and model templates with no clusters above $M_{200}(\rho_{\rm crit}) = 3 \times 10^{14}\ M_\odot\ h^{-1}$ (see Section 4.1.2 for details).

The results of the fit with no clusters masked are shown in Figure 1 and summarized in Table 2. The results of the fit with the two levels of cluster masking and mass function truncation are summarized in Tables 3 and 4. The best-fit parameter values using the nominal values of the beam and other sources of systematic uncertainty (see Section 4.2.1) are shown in the tables, along with 1σ statistical uncertainties (from the covariance matrix in the linear least-squares fit), 1σ systematic uncertainties (from the scatter in best-fit parameter values over 1000 realizations of systematic uncertainties), and the quadrature sum of the two 1σ uncertainties. The uncertainties on each parameter are the square root of the diagonal of the covariance matrix, i.e., the uncertainty of each individual parameter marginalized over the others. The parameter correlation matrix (statistical-only and statistical-plus-systematic) for the fit results with no cluster masking is shown in Table 5. Full statistical and

Table 5
Parameter Correlation Matrices for Multi-band Fits

Stat. only	tSZ	CIB	P95	P150	P220
tSZ	1.00	0.32	0.39	0.17	−0.28
CIB	0.32	1.00	0.12	−0.61	−0.88
P95	0.39	0.12	1.00	0.07	−0.11
P150	0.17	−0.61	0.07	1.00	0.54
P220	−0.28	−0.88	−0.11	0.54	1.00
Stat. + syst.	tSZ	CIB	P95	P150	P220
tSZ	1.00	0.35	0.46	0.37	0.03
CIB	0.35	1.00	0.16	−0.35	−0.43
P95	0.46	0.16	1.00	0.17	0.01
P150	0.37	−0.35	0.17	1.00	0.35
P220	0.03	−0.43	0.01	0.35	1.00

Notes. Correlation between parameters for the multi-band fit with no clusters masked. The top table shows the normalized elements of the parameter covariance matrix $r_{\psi\omega} = C_{\psi\omega}/\sqrt{C_{\psi\psi}C_{\omega\omega}}$ for statistical covariance only; the bottom table shows the same quantities for the sum of statistical and systematic covariance (see Section 4.2.1 for details on the calculation of systematic covariance). The parameter labels refer to thermal SZ amplitude, clustered CIB amplitude, and Poisson point-source amplitudes in each of the three bands.

systematic error covariance matrices are downloadable from the SPT Web site, as are the tSZ bispectrum templates and the 1000 beam realizations used in the fits.

5.2.1. Best-fit Thermal SZ Amplitudes

We discuss the cosmological implications of our tSZ bispectrum measurement in Section 6.3; here we briefly discuss the best-fit amplitudes at the three different masking levels, and we compare the best-fit amplitude with no masking to the measurement of the tSZ real-space three-point function (skewness) in ACT data from Wilson et al. (2012).

First, we note that the best-fit amplitudes at each masking level (no clusters masked, clusters above $8 \times 10^{14}\,M_\odot\,h^{-1}$ masked, clusters above $3 \times 10^{14}\,M_\odot\,h^{-1}$ masked) relative to the model prediction for that level of masking are statistically consistent with one another and indicate a lower tSZ bispectrum amplitude than predicted by the fiducial model. The implications of this result are discussed in Section 6.3. The model predicts a tSZ bispectrum amplitude at $l_1 = l_2 = l_3 = 3000$ and 152.8 GHz (the tSZ-weighted band center of the SPT 150 GHz band) of -9.8, -7.5, and $-2.4 \times 10^{-11}\,\mu\mathrm{K}^3$ for the three masking levels. The model prediction for $B_{\mathrm{tSZ}}(l_{\mathrm{rad}} = 3000)$ and 152.8 GHz is -11.1, -8.3, and $-2.6 \times 10^{-11}\,\mu\mathrm{K}^3$ for the three masking levels. The best-fit results from Tables 2–4 therefore translate to 152.8 GHz tSZ amplitudes of -5.2, -4.4, and $-1.7 \times 10^{-11}\,\mu\mathrm{K}^3$ at $l_1 = l_2 = l_3 = 3000$ and values of -5.9, -4.9, and $-1.8 \times 10^{-11}\,\mu\mathrm{K}^3$ for $B_{\mathrm{tSZ}}(l_{\mathrm{rad}} = 3000)$ for the three masking levels.

Roughly one-third of the total tSZ bispectrum is coming from clusters below the mass threshold used for the cosmological constraints in Reichardt et al. (2013), implying that cosmological constraints from the tSZ bispectrum do contain information beyond what is already measured using cluster counts. This would appear to be somewhat inconsistent with Figure 3 in B12, which shows that less than 10% of the tSZ skewness spectrum at $l = 3000$ is predicted to come from clusters with $M_{500}(\rho_{\mathrm{crit}}) < 2 \times 10^{14}\,M_\odot\,h^{-1}$ (roughly equivalent to the Reichardt et al. 2013 mass threshold of $M_{200}(\rho_{\mathrm{crit}}) = 3 \times 10^{14}\,M_\odot\,h^{-1}$). However, the contribution in mass and redshift to our measurement of the full 3D bispectrum is weighted slightly differently than the contribution to the skewness spectrum at $l = 3000$. When we calculate $B(\langle z, \rangle M_{200}(\rho_{\mathrm{crit}}))$ using $B_{\mathrm{tSZ}}(l_{\mathrm{rad}} = 3000)$ (which tracks the full 3D measured bispectrum very closely, see Section 6.3.3 for details), we find that the prediction is that roughly 25% of our measured signal should come from clusters below the Reichardt et al. (2013) mass threshold, consistent with what we observe. The general statement that the bispectrum is dominated by massive, low-redshift clusters still holds when $B_{\mathrm{tSZ}}(l_{\mathrm{rad}} = 3000)$ is used as the observable: in our model, 75% of the $B_{\mathrm{tSZ}}(l_{\mathrm{rad}} = 3000)$ signal is predicted to come from clusters with $M_{200}(\rho_{\mathrm{crit}}) > 3 \times 10^{14}\,M_\odot\,h^{-1}$ and $z < 0.6$.

To compare our Fourier-domain three-point function (i.e., bispectrum) tSZ amplitude to the real-space three-point function (skewness) of the tSZ measured in ACT data by Wilson et al. (2012), we first take the l-space filter shown in Figure 1 of Wilson et al. (2012) and calculate the real-space skewness that should be observed in a map convolved with this filter, if our tSZ bispectrum template is correct. To calculate this, we create a 3D bispectrum filter from the 1D Wilson et al. (2012) filter (by defining $F_{\mathrm{bisp}}(l_1, l_2, l_3) = F_{1d}(l_1)F_{1d}(l_2)F_{1d}(l_3)$), multiply the predicted tSZ bispectrum by this 3D filter, and calculate $\langle T^3 \rangle$ following Komatsu & Spergel (2001):

$$\left\langle T^3_{\mathrm{tSZ,filt}} \right\rangle = \frac{1}{2\pi^2} \sum_{l_1 l_2 l_3} (l_1 l_2 l_3) \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \times F_{\mathrm{bisp}}(l_1, l_2, l_3) B(l_1, l_2, l_3). \quad (32)$$

To evaluate the Wigner 3-j symbol, we use the high-l approximation in Equation (8) of B12.

The predicted skewness from our no-masking tSZ bispectrum template multiplied by the bispectrum version of the Wilson et al. (2012) filter is $-53.3\,\mu\mathrm{K}^3$ at 152.8 GHz. At the ACT band center of 148 GHz, the template and filter predict $-64.6\,\mu\mathrm{K}^3$. Given the amplitude we measure (in no-cluster-masked data) relative to the model prediction, and assuming the shape of the bispectrum model template is correct, our bispectrum measurement corresponds to a real-space tSZ skewness at 148 GHz of $-34.4 \pm 3.3\,\mu\mathrm{K}^3$, or $-34.4 \pm 7.7\,\mu\mathrm{K}^3$ when we add the 20% sample variance uncertainty estimated in Section 6.3.1. This is consistent with the value of $-31 \pm 14\,\mu\mathrm{K}^3$ (also including sample variance uncertainty) reported in Wilson et al. (2012).

Given the detection in this work of a significant clustered CIB bispectrum, it is likely that the Wilson et al. (2012) tSZ skewness measurement is biased low by ∼15% (see Section 6.2.1 for details). Wilson et al. (2012) recognized this potential bias. Their approach was to correct for it by measuring the CIB bispectrum in the Sehgal et al. (2010) simulations (scaled down by a factor of 1.7 in temperature) and subtracting that value from the tSZ skewness before using that number in cosmological fits. The correction was $-3.9 \pm 0.1\,\mu\mathrm{K}^3$, or 11% of the corrected tSZ skewness value of $-35 \pm 14\,\mu\mathrm{K}^3$. This bias estimate is roughly consistent with the prediction from the CIB bispectrum measured here, and the corrected Wilson et al. (2012) tSZ skewness is even more consistent than the uncorrected one with the tSZ bispectrum we measure.

5.2.2. χ^2 and Goodness-of-fit Values

The χ^2 values from the fits using the three levels of cluster masking and the nominal values of beams and other sources of systematic uncertainty are summarized in Table 6. The table

Table 6
χ^2 for Multi-band Fits

Cluster Masking	χ^2	χ^2, Null Hypothesis	$\Delta\chi^2$	$\chi^2_{\rm red}$
None	51798.1	52792.8	−994.8	1.10
$8 \times 10^{14}\ M_\odot\ h^{-1}$	51745.8	52788.5	−1042.7	1.10
$3 \times 10^{14}\ M_\odot\ h^{-1}$	50954.8	52063.7	−1108.9	1.08

Notes. χ^2 for simultaneous fits to bispectrum data in all three bands with three levels of cluster masking. Also shown is the χ^2 for the null hypothesis, the $\Delta\chi^2$ between the null hypothesis and the full model, and the reduced χ^2 for the full model (for 47,005 dof).

includes values of absolute χ^2, reduced χ^2, and $\Delta\chi^2$ between the best fit and the null hypothesis. The formal probabilities to exceed (PTEs) associated with the reduced χ^2 for all three levels of cluster masking are vanishingly small, but a small underestimate of the bispectrum variance would cause an otherwise good fit to have such a PTE. Since the χ^2 of the bispectrum fit will scale as the amplitude of the map noise and simulated Gaussian sky signal to the −6 power, the observed χ^2 excess is consistent with a percent-level misestimate in the noise or the Gaussian sky amplitude.

The map noise used to estimate the bispectrum variance is taken from the same data used to construct the real maps used to measure the bispectrum (see Section 3.4 for details). Thus, we can calculate a χ^2 from the "measured" bispectrum of every simulated map and use the scatter in χ^2 across the simulations as a measure of how closely the estimated bispectrum variance from map noise should match the map-noise variance contribution to the real data. None of the 100 simulations had χ^2 as high as the data, so it is unlikely that the excess χ^2 is due to a map noise misestimate. On the other hand, it is plausible that the Gaussian sky amplitudes could be mismatched between the simulations and the data at the 1% level. Our estimate for the amplitude of CMB fluctuations in SPT maps of this 800 deg^2 region is limited by cosmic variance and the uncertainty on our absolute calibration (which is 1%–2% in temperature, see Section 4.2.1), while our estimate for the Poisson point-source power is limited by calibration and beam uncertainties and by the lack of high-precision measurements of the Poisson amplitude at millimeter wavelengths.

Alternatively, the excess χ^2 could be evidence of departures from our models for either tSZ or clustered CIB. However, the total $\Delta\chi^2$ between the null hypothesis of zero bispectrum and the best-fit model is smaller than the difference between the χ^2 of the best-fit model and a χ^2 that would reduce to 1.0. Misestimates of the beam or filter transfer function could also be responsible. We can test this hypothesis by examining the χ^2 values for each systematic realization, and we do not find any trend of χ^2 with beam realizations; in fact, the total spread in χ^2 across all systematic realizations is roughly ±10, indicating that none of the identified sources of systematic uncertainty are responsible for the excess χ^2. Finally, the excess χ^2 is not strongly concentrated in one frequency band or region of l space. This points to a slight underestimate in the simulated Gaussian sky signal as the source of the excess χ^2.

6. DISCUSSION

In this section, we discuss the implications of each bispectrum component measured in the multi-band fit. We begin by comparing the amplitude of the Poisson point-source bispectrum in each band to model predictions; we then discuss the clustered CIB amplitude, both as an interesting signal in its own right and as a possible contaminant to the tSZ amplitude; finally, we implement the analysis introduced in B12 and use the tSZ bispectrum amplitude constraint to measure σ_8 and to sharpen the kSZ amplitude measurement from R12.

6.1. Poisson Point-source Component Amplitudes versus Model Predictions

Given a model for the number of sources in a given flux interval per unit solid angle $dN/dS/d\Omega$, we can predict the contribution to the bispectrum from the Poisson component of those sources. We can then compare these predictions to the results in Tables 2–4 as a test of the source models. Because the Poisson contribution to the bispectrum is weighted by the individual source fluxes cubed—compared to the source-flux-squared weighting in the power spectrum—this test is largely independent of power-spectrum-based tests of source models. And, because the bispectrum in this work is calculated with all sources detected above 5σ masked, the bispectrum constraints on models are nearly independent of source-count constraints from the same data (Vieira et al. 2010; Mocanu et al. 2013).

In Table 7, we show the predicted bispectrum amplitude in all three SPT bands from two models of radio-loud sources (De Zotti et al. 2010 and Tucci et al. 2011), two models of radio-quiet dusty sources (Béthermin et al. 2011 and Béthermin et al. 2012), and the four possible combinations of these models. We also repeat the measured values of the Poisson bispectrum component from Tables 2–4 for comparison. In some cases, the model predictions are at the nominal SPT bands, in others, the predictions are for the analogous *Planck* bands; in either case, we transform the models to the appropriate effective SPT band center for that source family, using assumed spectral indices of $\alpha_{\rm radio} = -0.5$ and $\alpha_{\rm dusty} = 3.5$, consistent with the results of Vieira et al. (2010) and R12. We also use this assumed spectral behavior to transform the 150 GHz flux cut to the other two bands.

Two things are immediately clear from Table 7. The first is that only in the 150 GHz band is the bispectrum expected to contain a significant contribution from both families of sources: at 95 GHz, the dusty sources are expected to contribute <5% of the total amplitude, while at 220 GHz, the radio-loud sources are expected to contribute <1% of the total amplitude. The second is that, while the De Zotti et al. (2010) model prediction is within 1σ of the measured 95 GHz measurement, there are significant differences between the model predictions and the measured amplitudes in all other cases.

We first investigate whether this discrepancy between measured and predicted Poisson bispectrum amplitudes could be due to effects that are not included in the measured uncertainty, in particular sample variance and the effect of a varying flux cut. Near the ∼6 mJy (at 150 GHz) flux cut used in this work, the dependence of radio source counts on flux is shallow (De Zotti et al. 2010; Vieira et al. 2010; Tucci et al. 2011; Mocanu et al. 2013). This means that the radio-source bispectrum will be dominated by the brightest (and rarest) sources just below the flux cut. This will tend to make the radio source bispectrum more sensitive to sample variance and to the fact that, while the real flux cut used in this work varies from field to field by ∼10%, we calculate the predicted bispectrum from source models using a single flux cut. We estimate the magnitude of both of these effects by simulated observations of many 800 deg^2 patches of sky containing sources drawn from the source count models listed in Table 7. In one set of simulated observations, we use the nominal 6 mJy 150 GHz source cut in every field;

Table 7
Poisson Source Component Results versus Model Predictions

Measured Poisson Bispectrum Amplitudes			
Masking level	95 GHz value ($10^{-10}\,\mu\mathrm{K}^3$)	150 GHz value ($10^{-11}\,\mu\mathrm{K}^3$)	220 GHz value ($10^{-10}\,\mu\mathrm{K}^3$)
No cluster masking	0.79 ± 0.16	1.12 ± 0.13	1.84 ± 0.35
$M_{200}(\rho_{\mathrm{crit}}) > 8 \times 10^{14}\,M_\odot\,h^{-1}$ clusters masked	0.88 ± 0.17	1.25 ± 0.14	1.73 ± 0.34
$M_{200}(\rho_{\mathrm{crit}}) > 3 \times 10^{14}\,M_\odot\,h^{-1}$ clusters masked	0.93 ± 0.17	1.27 ± 0.14	1.67 ± 0.33
Model Predictions			
Model	95 GHz value ($10^{-10}\,\mu\mathrm{K}^3$)	150 GHz value ($10^{-11}\,\mu\mathrm{K}^3$)	220 GHz value ($10^{-10}\,\mu\mathrm{K}^3$)
de Zotti radio	0.685	0.563	0.018
Tucci radio	0.484	0.385	0.011
Béthermin et al. (2011) dusty	0.023	3.285	2.264
Béthermin et al. (2012) dusty	0.016	1.878	2.900
de Zotti et al. + Béthermin et al. (2011)	0.708	3.849	2.282
de Zotti et al. + Béthermin et al. (2012)	0.701	2.441	2.918
Tucci et al. + Béthermin et al. (2011)	0.507	3.670	2.275
Tucci et al. + Béthermin et al. (2012)	0.500	2.263	2.911

Notes. Comparison of measured single-band Poisson point-source bispectrum amplitudes with predictions from source count models. In the upper section, measured values of the Poisson source-component bispectrum amplitudes—and 1σ statistical-plus-systematic errors on those values—are shown for the three levels of cluster masking. In the lower section, predicted Poisson bispectrum amplitudes are shown for two models of radio-loud source counts (De Zotti et al. 2010 and Tucci et al. 2011), two models of dusty, radio-quiet source counts (Béthermin et al. 2011 and Béthermin et al. 2012), and combinations thereof.

in another set, we use the actual 150 GHz source cut used in each individual field in this work; these cut levels range from 5.7 to 6.6 mJy. Both the bispectrum sample variance (calculated as the scatter among the best-fit Poisson bispectrum in all simulated observations) and the effect of the different flux cuts were largest at 95 GHz—not surprising, given that this is the band in which the radio source contribution is largest—but even in that band, the square root of the sample variance was only 2% of the average Poisson bispectrum, and the difference between using the true flux cut for every field and using the nominal flux cut was only 6%. The effect of sample variance on the bispectrum due to dusty sources should be significantly smaller than this, because the predicted dusty source bispectrum is dominated by sources well below the flux cut used here. We conclude that the discrepancy between model predictions and the measured Poisson bispectrum cannot be explained by sample variance and varying flux cuts.

The simplest modifications to the source models that would bring them in line with the bispectrum results in this work would be: (1) to steepen the spectral behavior used to extrapolate the dusty source behavior from higher frequencies to the SPT bands, thus reducing the dusty-source bispectrum by a small amount at 220 GHz and a larger amount at 150 GHz; and (2) to assume a slightly shallower spectral index in extrapolating counts at radio frequencies to 95 GHz, particularly for the Tucci et al. (2011) model. It remains to be seen whether such modifications would be compatible with constraints from other measures of point-source behavior such as number counts and the power spectrum. An interesting possibility for future work is to combine these probes into a simultaneous constraint on source models.

6.2. The Clustered CIB Bispectrum

Measurements of the two-point function of CIB clustering (either the real-space two-point angular correlation function or the angular power spectrum) are currently providing key constraints on the relationship between star-forming galaxies and their dark-matter halos, or, equivalently, on the relationship between luminosity and mass in star-forming galaxies (see, e.g., Viero et al. 2013 and references therein). Equally interesting will be constraints on this relationship from the clustered CIB bispectrum. As is the case for tSZ and the Poisson point-source component, the relative weighting of sources that contribute to the clustered CIB power spectrum and bispectrum will be different—with the bispectrum generally sensitive to brighter sources because of the S^3 weighting—implying that the two probes can give independent constraints on models of the mass–luminosity relationship.

There are currently no physically motivated predictions for the clustered CIB bispectrum in the literature—although Lacasa (2012) present a heuristic model based on power spectrum measurements, which we adopt as our fitting template. However, any model of the mass–luminosity relationship of star-forming galaxies that can predict the clustered CIB power spectrum should also be able to predict the bispectrum, and we expect the measurement of the clustered CIB bispectrum in this work to provide new constraints on such models.

For now, our main conclusions about the clustered CIB bispectrum are that it is clearly detectable in 800 deg^2 of 220 GHz data at SPT noise levels, and that the angular shape of the signal is fit reasonably well by a pure power law. Our model, based on the ansatz of Lacasa (2012) and described in Section 4.1.3, has $B_{\mathrm{clust}}(l_1, l_2, l_3) \propto (l_1 l_2 l_3)^{-n/2}$, with $n = 1.2$ and scales with observing frequency as $\nu^{3\alpha}$, with $\alpha = 3.72$. As shown in Figure 1, this is by eye a reasonable fit to the data. Although the PTE associated with the reduced χ^2 is vanishingly small, as discussed in Section 5.2.2, this is consistent with a very small noise misestimate. Neither changing the power-law index of the angular shape of the signal nor changing the assumed frequency scaling results in significant improvements in χ^2 ($|\Delta\chi^2| < 2$ for a 2σ shift in α or a 30% change in the power-law index). For the fiducial model, using the results of the no-cluster-masking multi-band fit, the amplitude of the clustered CIB bispectrum at $l_1 = l_2 = l_3 = 3000$ and 219.6 GHz (the CIB-weighted band center of the SPT 220 GHz band) is

$(3.21 \pm 0.68) \times 10^{-10}\ \mu\mathrm{K}^3$. Using the $>8 \times 10^{14}\ M_\odot\ h^{-1}$ masking result, the amplitude of the clustered CIB bispectrum at $l_1 = l_2 = l_3 = 3000$ and 219.6 GHz is $(3.51 \pm 0.70) \times 10^{-10}\ \mu\mathrm{K}^3$.

6.2.1. The Clustered CIB as a Contaminant to the Thermal SZ Bispectrum

The clustered CIB bispectrum that we detect in the SPT 220 GHz band will also contribute to the bispectrum at 150 GHz (and, to a much lesser extent, at 95 GHz). With the assumed frequency scaling of $B \propto \nu^{3\alpha}$, with $\alpha = 3.72$, the ratio of clustered CIB bispectrum amplitude in the 150 and 95 GHz bands compared to that in the 220 GHz band should be 0.031 and 0.001, respectively, meaning we would expect roughly $1 \times 10^{-11}\ \mu\mathrm{K}^3$ of clustered CIB bispectrum at $l = 3000$ and 150 GHz (compared to an expected tSZ bispectrum of $-5.4 \times 10^{-11}\ \mu\mathrm{K}^3$) and a negligible contribution of $<10^{-12}\ \mu\mathrm{K}^3$ at 95 GHz. This implies that, if we were to neglect the clustered CIB, we would underestimate the tSZ bispectrum amplitude by roughly 20% at 150 GHz (because the bispectrum shape of the tSZ and clustered CIB are similar, but with different polarities). If we fit both the 95 and 150 GHz bispectra individually to a two-component model consisting of tSZ and a Poisson point-source term, the results are as expected. The best-fit tSZ amplitude with no clusters cut from the 95 GHz data alone ($B_{\mathrm{tSZ}} = 0.54 \pm 0.07$) is consistent with the multi-band fit results ($B_{\mathrm{tSZ}} = 0.54 \pm 0.04$, see Table 2), but the best-fit tSZ amplitude from the 150 GHz data alone ($B_{\mathrm{tSZ}} = 0.43 \pm 0.05$) is ∼20% lower than the multi-band result.

The multi-band fit properly accounts for this, and if the CIB behavior were very different from what we assume in the model, this would manifest in a noticeably poorer χ^2 in the multi-band fit relative to single-band fits, which we do not see. In particular, if there were a significant level of tSZ–CIB correlation in the bispectrum, we would expect to see a very different best-fit tSZ amplitude from the multi-band fit from what we obtain with the 95 GHz-only fit; in fact, our cosmological constraints detailed below would not substantively change if we used only 95 GHz data in the fit, although the error bars would increase somewhat. We conclude that, at the current level of statistical precision, we see no evidence that our model of the CIB is inadequate or that the CIB is significantly biasing our measurement of the tSZ bispectrum. More complicated models involving spatial correlations between the sources of the tSZ and CIB bispectra, different CIB bispectrum shapes, and spectral behavior beyond a single spectral index will be explored in future analyses which include measurements of the cross-bispectra among the three SPT bands (in addition to the auto-bispectra considered here).

As noted in Section 5.2.1, according to our model and fit results, the 148 GHz tSZ skewness measurement of Wilson et al. (2012) in ACT data is likely to be biased low by roughly 15% (less than the 20% we see at 152.6 GHz, the CIB-weighted band center of the SPT 150 GHz band, due to the very steep frequency scaling of the CIB). This is slightly smaller than the statistical + sample variance uncertainty in that result—and significantly smaller than the difference in tSZ skewness predicted by the range of ICM models they consider; we also note that Wilson et al. (2012) included an 11% correction for CIB contamination in the tSZ skewness value they used in cosmological fits.

6.3. Cosmological Interpretation of the Thermal SZ Bispectrum Amplitude

In this section, we use our measurement of B_{tSZ}, the tSZ bispectrum amplitude, to place a constraint on σ_8 and to predict the tSZ power spectrum amplitude, A_{tSZ}. We use this A_{tSZ} prediction to break degeneracies between tSZ and kSZ in measurements of the CMB power spectrum. First, however, we need to estimate two properties of the A_{tSZ} and B_{tSZ} distributions, namely the sample variance of B_{tSZ} and the covariance of A_{tSZ} with B_{tSZ} over the same patch of sky.

6.3.1. Sample Variance in the Measurement of B_{tSZ}

To estimate the sample variance contribution to our measurement of B_{tSZ}, we use a set of cosmological simulations. These simulations use the same gas physics prescription, gas physics parameter settings, and cosmological parameter settings that went into the template predictions used in the model fitting procedure described in Section 4.2. The simulations cover an octant of sky, from which we extract 40 independent 100 deg^2 fields. We run the bispectrum estimator over these fields with the same weighting used in running the estimator on the 150 GHz data. We fit each of the resulting 40 bispectrum measurements to the predicted template, again using the weights from the 150 GHz data and restricting the fit to $l \leqslant 4000$ to roughly account for the effects of the Poisson point-source component in the fit to the data. The calculation is performed with 10 levels of cluster masking, ranging from no masking to masking clusters above $M_{200}(\rho_{\mathrm{crit}}) \geqslant 2 \times 10^{14}\ M_\odot\ h^{-1}$.

We then estimate the scatter in 800 deg^2 regions for each cluster masking level by grouping the 40 amplitudes into five independent groups of eight amplitudes, averaging each group, and calculating the scatter among groups. This is a noisy estimate of the sample variance. In particular, the sample variance as a function of mask threshold is affected by the masking of individual clusters at high enough mass that only a few such clusters exist in the entire octant simulation. For this reason, we fit a smooth function to the measured sample variance as a function of masking, and we report our sample variance as the best-fit value at the three masking levels used for the data. The fractional scatter in B_{tSZ} due to sample variance is estimated to be 0.20, 0.15, and 0.06 at the three levels of cluster masking (none, $>8 \times 10^{14}\ M_\odot\ h^{-1}$, $>3 \times 10^{14}\ M_\odot\ h^{-1}$) used for the data. We note that the value for the no-masking case is consistent with the sample variance of the tSZ skewness measured by Wilson et al. (2012), given the relative sky coverage of the two analyses. (Wilson et al. 2012 measured 41% scatter for 239 deg^2 as compared to 20% for 837 deg^2 in this work.)

There is also a potential systematic uncertainty in B_{tSZ} introduced when clusters are masked. If the mass estimates for all clusters are systematically biased, then the mask threshold used in the model to which the data is compared is different than the mask threshold actually used in the data. The uncertainty on the overall scaling between the SPT measure of SZ signal and cluster mass, as estimated in Benson et al. (2013), is 10% at $z = 0$ and 15% at $z = 1$. This includes the contribution from the uncertainty in the weak-lensing-derived normalization of the X-ray Y_X–mass relation. Although the bispectrum is dominated by low-redshift clusters, we adopt the 15% uncertainty to be conservative. Using our model for the tSZ bispectrum—specifically $B_{\mathrm{tSZ}}(l_{\mathrm{rad}} = 3000)$—as a function of mass and redshift, we estimate the effect of the systematic mass uncertainty on our determination of B_{tSZ} by integrating the model prediction over redshift and up to three different maximum mass values: the nominal value we use, and that value times 1.15 and 0.85. We find that a 15% error in mass leads to a ∼10% error in B_{tSZ} for our $M_{200}(\rho_{\mathrm{crit}}) = 8 \times 10^{14}\ M_\odot\ h^{-1}$ mass cut and a ∼30%

Table 8
Thermal SZ Bispectrum Error Budget

Masking	Stat. Error	Syst. Error	Mask Thresh. Error	Sample-variance Error	Stat. + Syst. + Mask+ Sample-variance
No cluster masking	0.08	0.06	0.00	0.20	0.22
$M_{200}(\rho_{\rm crit}) > 8 \times 10^{14}\, M_\odot\, h^{-1}$ clusters masked	0.09	0.06	0.11	0.15	0.21
$M_{200}(\rho_{\rm crit}) > 3 \times 10^{14}\, M_\odot\, h^{-1}$ clusters masked	0.25	0.09	0.32	0.06	0.42

Notes. Fractional 1σ uncertainty on the amplitude of the tSZ bispectrum, including statistical, systematic, cluster-mask-threshold, and sample variance contributions, and the combination of these in quadrature. Sample variance errors are estimated as described in Section 6.3.1. Errors due to the uncertainty in cluster mask threshold are calculated assuming a 15% systematic uncertainty in cluster mass estimation (see Section 4.1.2 for details). Values are given for each of the three cluster mask thresholds (no masking, clusters with $M_{200}(\rho_{\rm crit}) > 8 \times 10^{14}\, M_\odot\, h^{-1}$ masked, clusters with $M_{200}(\rho_{\rm crit}) > 3 \times 10^{14}\, M_\odot\, h^{-1}$ masked).

error for our $M_{200}(\rho_{\rm crit}) = 3 \times 10^{14}\, M_\odot\, h^{-1}$ cut. We add this to the total error budget on $B_{\rm tSZ}$ in all calculations that follow.

As discussed in Section 4.1.2, the sample variance and statistical-plus-systematic-plus-mask-threshold uncertainties on $B_{\rm tSZ}$ go in opposite directions as more clusters are masked, implying that there is an optimal mass threshold, at which level the total error on $B_{\rm tSZ}$ is minimized. Table 8 shows the fractional uncertainty from each of these sources (and the quadrature sum of all of them) for the three masking levels used for the data. Among these three masking levels, the total uncertainty is smallest when clusters above $8 \times 10^{14}\, M_\odot\, h^{-1}$ are masked.

To investigate whether a different threshold would be optimal, we calculate the total uncertainty at the 10 mask thresholds used in the sample variance calculation. At each mask threshold, we re-calculate the uncertainty in $B_{\rm tSZ}$ due to cluster mass systematic error, and we scale the fractional statistical error by the best-fit $B_{\rm tSZ}$ in the simulations using that mask threshold. This calculation implies that the total fractional scatter has a broad minimum between $6 \times 10^{14}\, M_\odot\, h^{-1}$ and $9 \times 10^{14}\, M_\odot\, h^{-1}$. We use the $8 \times 10^{14}\, M_\odot\, h^{-1}$ cut values of $B_{\rm tSZ}$ for our cosmological results.

6.3.2. Covariance between A_{tSZ} and B_{tSZ}

We estimate the covariance between $A_{\rm tSZ}$ and $B_{\rm tSZ}$ using the halo-model approach of Kayo et al. (2013), together with the gas physics prescription of B12. We find that the square root of the fractional covariance between $A_{\rm tSZ}$ with no clusters cut and $B_{\rm tSZ}$ with clusters above $8 \times 10^{14}\, M_\odot\, h^{-1}$ cut is $\sim$0.06. This is small compared to the other sources of uncertainty in our prediction of $A_{\rm tSZ}$, and we ignore it for this analysis. Details of the $A_{\rm tSZ}/B_{\rm tSZ}$ covariance calculation are given in the Appendix.

6.3.3. A σ_8 Constraint from the Thermal SZ Bispectrum

In this section, we translate our measurement of the amplitude of the tSZ bispectrum into a constraint on σ_8. Section 4.1.2, summarizing B12, describes our model for the tSZ bispectrum and its dependence on cosmological parameters. We use this model to determine the cosmological scaling of the tSZ bispectrum amplitude as well as the modeling uncertainty associated with our measurement. Rather than replicating the full 3D fit as it is performed for the data, we work with the 1D quantity $B_{\rm tSZ}(l_{\rm rad} = 3000)$ when determining the cosmological scaling and modeling uncertainty. This allows for a more straightforward generalization to other data sets and experiments. Note that B12 employed a similar approach but used the value of the tSZ skewness spectrum at $l = 3000$ rather than $B_{\rm tSZ}(l_{\rm rad} = 3000)$. The choice of $B_{\rm tSZ}(l_{\rm rad} = 3000)$ as a proxy for the amplitude estimated from the full 3D fit is supported by tests with simulated data showing that the two observables track each other with less than 2% scatter in their ratio.

The modeling uncertainty and cosmological scaling calculated here for $B_{\rm tSZ}$ are slightly different than those quoted in B12. This is because a different proxy observable is used, and the most massive clusters are not included in the analysis presented here. We find a modeling uncertainty of 36%, compared to the 33% in B12 for the skewness spectrum at $l = 3000$ and no cluster cut. For a six-parameter flat ΛCDM model, we find the cosmological scaling of $B_{\rm tSZ}(l_{\rm rad} = 3000)$ with clusters above $8 \times 10^{14}\, M_\odot\, h^{-1}$ cut to be

$$B_{\rm tSZ}(l_{\rm rad} = 3000;\ M_{200}(\rho_{\rm crit}) < 8 \times 10^{14}\, M_\odot\, h^{-1}) \propto \left(\frac{\sigma_8}{0.8}\right)^{9.1} \left(\frac{\Omega_b}{0.045}\right)^{3.82} \left(\frac{h}{0.71}\right)^{2.25} \times \left(\frac{n_s}{0.97}\right)^{-1.12} \left(\frac{\Omega_m}{0.27}\right)^{-0.27}, \quad (33)$$

(with no measurable dependence on τ). Compared to the scaling in B12 (for the skewness spectrum at $l = 3000$ and no cluster cut), the primary difference is a slightly shallower scaling with σ_8($B_{\rm tSZ} \propto \sigma_8^{11.6}$ in B12).

To compare our result to the model predictions, we first translate our best-fit amplitude with respect to the model prediction for the tSZ bispectrum (presented in Section 5.2 and Tables 2–4) into a value of $B_{\rm tSZ}(l_{\rm rad} = 3000)$ by multiplying the model by our best-fit amplitude and summing the model as in Equation (30), using the weights from the 150 GHz data. We use the result from our fit with all clusters above $M_{200}(\rho_{\rm crit}) = 8 \times 10^{14}\, M_\odot\, h^{-1}$ masked, and we use the mask threshold error, sample variance, and modeling uncertainty estimated for that mass cut. We marginalize over all cosmological parameters other than σ_8. Although the dependence of $B_{\rm tSZ}$ on σ_8 is far stronger than on the other parameters, the dependence on Ω_b is strong enough that we place a *WMAP7* (Larson et al. 2011) prior on $\Omega_b h^2$ and a prior on h from Riess et al. (2011).

Taking into account the full uncertainty (statistical + systematic + mask threshold + sample variance) on our measurement, and adding the 36% modeling uncertainty, the resulting constraint on σ_8 is

$$\sigma_8 = 0.787 \pm 0.031. \quad (34)$$

The uncertainty on our determination of σ_8 is dominated by the assumed 36% modeling uncertainty. Given the steep scaling of σ_8 with $B_{\rm tSZ}$ and the mild dependence on other parameters, in the absence of modeling uncertainty, we would expect to

achieve a ∼2% constraint on σ_8 from our 21% constraint on $B_{\rm tSZ}$, compared to the 4% constraint achieved when modeling uncertainty is included.

This constraint on σ_8 from the tSZ bispectrum is comparable in significance to, and statistically consistent with, other recent determinations of σ_8 from tSZ and/or the primary CMB. From the primary CMB, in a flat ΛCDM model, Hinshaw et al. (2013) find $\sigma_8 = 0.821 \pm 0.023$ from *WMAP9* data alone, while Story et al. (2013) find $\sigma_8 = 0.795 \pm 0.022$ when adding SPT constraints from the primary CMB damping tail to *WMAP7* data. Adding constraints from the tSZ power spectrum to *WMAP7* plus earlier damping-tail constraints, Shirokoff et al. (2011) obtain a constraint on σ_8 with statistical precision at the ±0.01 level, but which varies in best-fit value from $0.77 < \sigma_8 < 0.80$ depending on the model template used. Similarly, Dunkley et al. (2011), using only tSZ power spectrum data, obtain a ±0.05 constraint (statistical only) but find best-fit values from $0.74 < \sigma_8 < 0.79$, depending on the choice of model template. Adding SPT cluster counts to *WMAP7* and the Keisler et al. (2011) SPT measurement of the damping tail, and marginalizing over uncertainty in the X-ray-calibrated tSZ–mass scaling relation from Benson et al. (2013), Reichardt et al. (2013) find $\sigma_8 = 0.798 \pm 0.017$. Combining *WMAP7* with ACT-detected clusters and marginalizing over uncertainty in a dynamical-mass-calibrated scaling relation, Hasselfield et al. (2013) find $\sigma_8 = 0.829 \pm 0.024$.

Finally, in the analysis most directly comparable to the one presented here, Wilson et al. (2012) find $\sigma_8 = 0.79 \pm 0.03$ from a measurement of the tSZ real-space three-point function (skewness). Wilson et al. (2012) do not explicitly marginalize over modeling uncertainty, but they obtain σ_8 constraints using different gas model prescriptions and find that, even for the extreme case of turning off cooling, feedback, and star formation, the constraint on σ_8 changes by less than 1σ. The agreement between the SPT and ACT constraints on σ_8 from the tSZ three-point function is not surprising, given the consistency between the measured Fourier-domain and real-space three-point amplitudes discussed in Section 5.2.1. This consistency is worthy of note, however, given the very different analysis techniques leading to the two constraints and the different regions of the sky measured.

6.3.4. Predicting A_{tSZ} and Sharpening A_{kSZ}

Using the approach of B12, we now convert our constraint on the amplitude of the tSZ bispectrum into a prediction for $A_{\rm tSZ}$, the amplitude of the tSZ power spectrum. We then use that prediction to sharpen the R12 constraint on $A_{\rm kSZ}$, the amplitude of the kSZ power spectrum.

We do not apply any cluster cut to the bispectrum-derived prediction for $A_{\rm tSZ}$ or to the measurement of $A_{\rm tSZ}$ from SPT power spectrum data (which we take directly from R12). As detailed in B12 (and references therein), the tSZ power spectrum is dominated by lower-mass halos, so the mass-function uncertainties at the high-mass end are not as important for the tSZ power spectrum as they are for the tSZ bispectrum.

As in the previous section, we use the value of $B_{\rm tSZ}(l_{\rm rad} = 3000)$ as a proxy for the results of the full, 3D fit of our data to the model predictions for $B_{\rm tSZ}(l_1, l_2, l_3)$. We find a slightly different scaling between $A_{\rm tSZ}$ and $B_{\rm tSZ}(l_{\rm rad} = 3000)$ with clusters above $8 \times 10^{14}\, M_\odot\, h^{-1}$ cut than between $A_{\rm tSZ}$(no cut) and the tSZ skewness spectrum at $l = 3000$ and no clusters cut. Specifically, B12 found

$$B_{\rm tSZ} \propto A_{\rm tSZ}^{1.4} \tag{35}$$

using the skewness spectrum at $l = 3000$ and no clusters cut, while we find

$$B_{\rm tSZ} \propto A_{\rm tSZ}^{1.14} \tag{36}$$

using $B_{\rm tSZ}(l_{\rm rad} = 3000)$ and clusters above $8 \times 10^{14}\, M_\odot\, h^{-1}$ cut. We also find a slightly different uncertainty in our prediction of $A_{\rm tSZ}$ given $B_{\rm tSZ}$, namely 11% for the default B12 gas physics assumptions and 18% for the extreme case (as compared to 7% and 15% for $A_{\rm tSZ}$ at fixed $B_{\rm tSZ}$ using the skewness spectrum at $l = 3000$ and no cluster cut).

As detailed in B12, the bispectrum measurement acts to constrain gas model parameters (including redshift evolution, which is poorly constrained without the bispectrum measurement). The effects of the gas model on predictions of $A_{\rm tSZ}$ and $B_{\rm tSZ}$ are highly correlated, so a bispectrum measurement allows us to reduce the uncertainty on $A_{\rm tSZ}$ significantly.

R12 report $A_{\rm tSZ}$ and $A_{\rm kSZ}$ in terms of the power—when expressed as $D_l = l(l+1)C_l/2\pi$—at $l = 3000$. Using the best-fit bispectrum tSZ amplitude measurement with clusters above $8 \times 10^{14}\, M_\odot\, h^{-1}$ masked (including statistical, systematic, and sample-variance uncertainties), the prediction for $A_{\rm tSZ}$ using the default B12 11% modeling uncertainty is

$$A_{\rm tSZ}(\text{predicted from } B_{\rm tSZ}) = 2.96 \pm 0.64\, \mu{\rm K}^2. \tag{37}$$

Increasing the modeling uncertainty to 18% increases this error bar to $\pm 0.77\, \mu{\rm K}^2$.

We create a Gaussian prior from this prediction and use it to importance sample the posterior probability distributions from R12. In that work, $A_{\rm tSZ}$ and $A_{\rm kSZ}$ were estimated using two different assumptions about the spatial correlation between tSZ and CIB: (1) assuming zero correlation, and (2) assuming a single correlation coefficient independent of angular scale and allowing that coefficient to be a free parameter. If we importance-sample the zero-correlation result from R12, the improvement from the bispectrum prior is small. In the case of tSZ–CIB correlation as a free parameter, however, the bispectrum prior reduces the tSZ uncertainty by nearly a factor of two and results in a posterior $A_{\rm kSZ}$ distribution with a clear non-zero peak (see Figure 2, right panel). There is some modeling inconsistency in using the $B_{\rm tSZ}$ constraint from this work, which is derived assuming no tSZ–CIB correlation, to improve the R12 $A_{\rm tSZ}$ constraint derived with tSZ–CIB correlation as a free parameter. However, as detailed in Section 4.1.3, we expect tSZ–CIB correlation to affect the bispectrum far less than the power spectrum, such that we can ignore the effects of tSZ–CIB correlation on our measurement of $B_{\rm tSZ}$.

The bispectrum-informed constraints on $A_{\rm tSZ}$ and $A_{\rm kSZ}$ from R12, with tSZ–CIB correlation as a free parameter (and using the Shaw et al. 2012 cooling + star formation template for kSZ), are

$$\begin{aligned} A_{\rm tSZ} &= 3.08 \pm 0.56\, \mu{\rm K}^2 \\ A_{\rm kSZ} &= 2.9 \pm 1.6\, \mu{\rm K}^2 \\ A_{\rm kSZ} &< 5.6\, \mu{\rm K}^2\ (95\%). \end{aligned} \tag{38}$$

These results are fairly insensitive to modeling uncertainty: using the extreme 18% modeling uncertainty instead of the default 11% value increases the 1σ error on $A_{\rm tSZ}$ to $\pm 0.63\, \mu{\rm K}^2$, while the 1σ error on $A_{\rm kSZ}$ is effectively unchanged. We note that this result is lower in power than the best-fit $A_{\rm kSZ}$ of $5.3^{+2.2}_{-2.4}\, \mu{\rm K}^2$ found by Addison et al. (2013) using combined SPT, ACT, *Herschel*-SPIRE, and *Planck* data, but that the two results are consistent at the 1σ level.

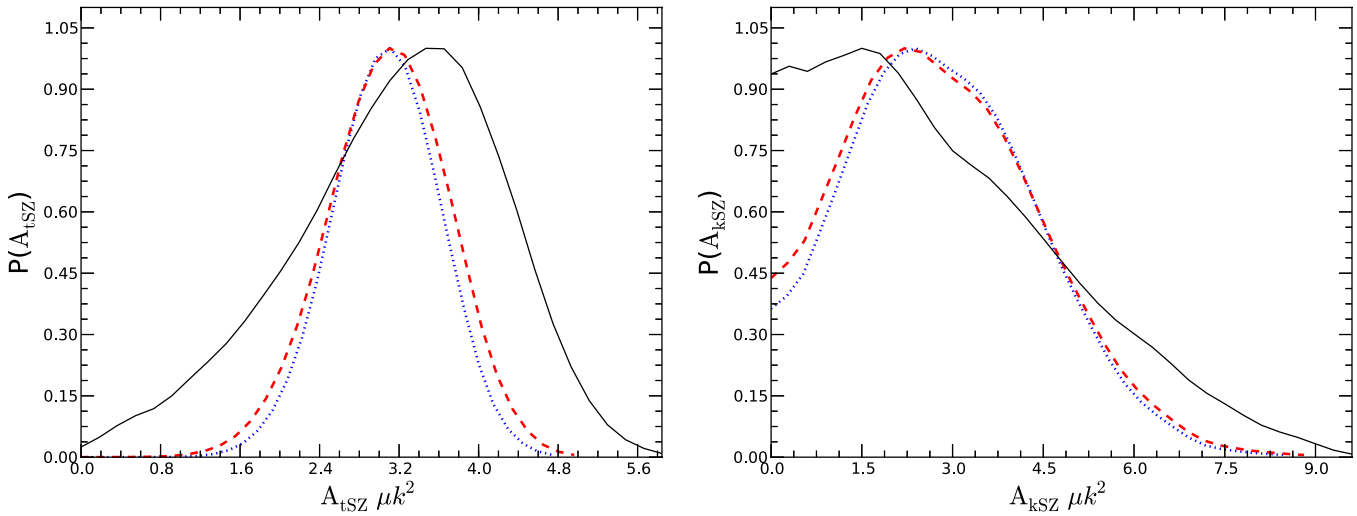


Figure 2. One-dimensional posterior probability distributions from R12 for $A_{\rm tSZ}$ (left) and $A_{\rm kSZ}$ (right), in the case in which the spatial correlation between tSZ and CIB was a free parameter in the R12 fits, before and after applying the bispectrum-based prior in Equation (37). The black (solid) line shows constraints with no bispectrum information added; the blue (dotted) line shows the constraints assuming the default 11% modeling uncertainty in $A_{\rm tSZ}$ for fixed $B_{\rm tSZ}$; the red (dashed) line shows constraints assuming the extreme 18% modeling uncertainty. In all cases, adding constraints from the bispectrum data improves the power-spectrum-only constraints.

(A color version of this figure is available in the online journal.)

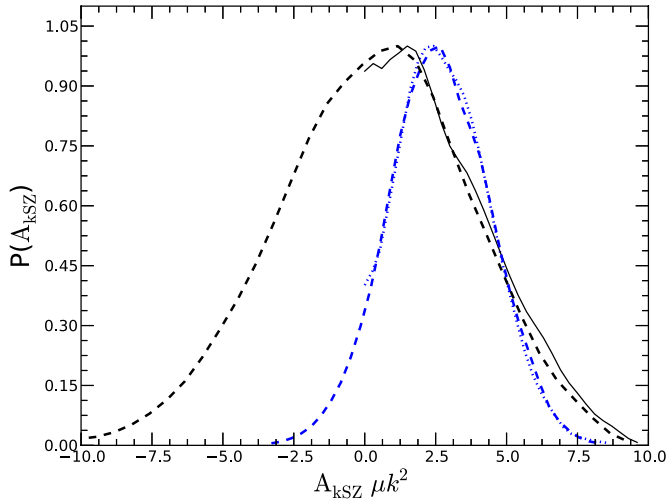


Figure 3. One-dimensional posterior probability distributions from R12 for $A_{\rm kSZ}$, with and without including bispectrum information, and with and without applying the $A_{\rm kSZ} > 0$ prior. The solid and dashed black lines show constraints with no bispectrum information added, with and without the $A_{\rm kSZ} > 0$ prior. (The black solid line is identical to the black solid line in the right panel of Figure 2.) The dotted and dashed blue lines show the constraints with bispectrum information added, with and without the $A_{\rm kSZ} > 0$ prior. (The blue dotted line is identical to blue dotted line in the right panel of Figure 2.)

(A color version of this figure is available in the online journal.)

Note that we have applied a prior of $A_{\rm kSZ} > 0$ in obtaining these results. Under the assumption that we are measuring real sky power in the power spectrum data in R12, there is no reason to lift the $A_{\rm kSZ} > 0$ prior, but one could choose to account for possible unknown systematics by expanding the prior below zero. The $A_{\rm kSZ}$ posterior with no prior on $A_{\rm kSZ}$ is plotted (with and without the bispectrum information included) in Figure 3. The no-prior mean and 1σ values are $A_{\rm kSZ} = 2.6 \pm 1.8\,\mu{\rm K}^2$ when the bispectrum information is included, as compared to $A_{\rm kSZ} = 0.3 \pm 3.3\,\mu{\rm K}^2$ from the R12 power spectrum only.

Before adding the bispectrum constraint, the R12 1σ uncertainty on $A_{\rm tSZ}$ was $1.05\,\mu{\rm K}^2$, the 95% upper limit on $A_{\rm kSZ}$ was $6.7\,\mu{\rm K}^2$, and the peak of the $A_{\rm kSZ}$ distribution was within 1σ of zero. The addition of the bispectrum constraint reduces the error of $A_{\rm tSZ}$ by a factor of two compared to the power spectrum constraints alone. In turn, this improves the constraints on $A_{\rm kSZ}$, reducing the upper limit by 20% (and the 1σ uncertainty by nearly a factor of two in the no-kSZ-prior case), and showing a preference for non-zero kSZ.

6.3.5. Prospects for the Full 2500 deg^2 Survey

The full SPT-SZ survey comprises 2500 deg^2 of 95, 150, and 220 GHz data at noise levels comparable to the 800 deg^2 subset used in this analysis, and work is ongoing to produce the data and simulation products necessary to measure the small-angular-scale power spectrum and bispectrum in the full survey. The statistical uncertainty and the sample-variance uncertainty on $B_{\rm tSZ}$ from the full survey should be roughly a factor of $\sqrt{3}$ lower than the corresponding values in this work, simply from the larger sky coverage. The systematic uncertainty is not expected to change, but in the 800 deg^2 result, the statistical + systematic + sample variance uncertainty is dominated by the sample variance contribution in both the no-cluster-masking and $>8 \times 10^{14}\,M_\odot\,h^{-1}$ masking cases; this total uncertainty is also expected to decrease by nearly $\sqrt{3}$. For the $>8 \times 10^{14}\,M_\odot\,h^{-1}$ masking case, this would result in a $\sim$12% constraint on $B_{\rm tSZ}$.

Because the constraint on σ_8 from $B_{\rm tSZ}$ is already limited by the assumed 36% modeling uncertainty, we do not expect a measurable improvement in the σ_8 constraints from the full 2500 deg^2 survey, unless significant progress is made in measuring pressure profiles of the clusters responsible for the tSZ bispectrum. To achieve a lower statistical + systematic + sample variance uncertainty using the 2500 deg^2 result, we would need to reduce the modeling uncertainty by roughly a factor of two. This is an ambitious goal; however, the amount of X-ray and millimeter-wave data on high-mass clusters at all redshifts is increasing rapidly, with X-ray programs such as *Chandra* observations of 80 SPT-discovered clusters at $0.4 \leqslant z \leqslant 1.2$ (B. Benson et al., in preparation) and millimeter-wave pressure profile measurements from such instruments

as Bolocam (Sayers et al. 2013), the Combined Array for Research in Millimeter-wave Astronomy (e.g., Plagge et al. 2013), and *Planck* (Planck Collaboration et al. 2013).

Even with no improvement in modeling uncertainty, the 2500 deg^2 measurement of $B_{\rm tSZ}$ will improve our ability to separate $A_{\rm tSZ}$ and $A_{\rm kSZ}$ in the power spectrum. Because the relationship of $A_{\rm tSZ}$ to $B_{\rm tSZ}$ is constrained far better than either one individually, our current bispectrum-derived constraint on $A_{\rm tSZ}$ is limited by sample variance. Reducing the full statistical + systematic + sample variance + cluster mask threshold uncertainty on $B_{\rm tSZ}$ to $\sim$12% will result in uncertainties of $\sim$0.4 μK^2 on $A_{\rm tSZ}$ and $\sim$1.0 μK^2 on $A_{\rm kSZ}$, assuming the default 11% modeling uncertainty. Achieving this total error budget will also require an improvement in the systematic uncertainty on our cluster mass determinations, but that is expected to be achieved with a program of multi-wavelength follow-up of SPT clusters that is currently underway (see Benson et al. 2013 for details).

If the current best-fit value of $A_{\rm kSZ}$ turns out to be correct, these constraints will result in nearly a 3σ detection of the kSZ effect. The addition of 100 deg^2 of already collected *Herschel*-SPIRE submillimeter data (program OT1_jcarls01_3, PI: Carlstrom) will provide strong constraints on the behavior of the CIB, which in turn will further tighten the tSZ and kSZ constraints. Full-survey SPT power spectrum + full-survey SPT bispectrum + 100 deg^2 SPIRE constraints on the kSZ are expected to be at the $<$0.5 μK^2 level. These constraints will lead to unprecedented limits on the reionization history of the universe (e.g., Zahn et al. 2012).

7. CONCLUSIONS

We have used 800 deg^2 of multi-frequency data from the SPT-SZ survey to make a high-significance detection of the Fourier-domain angular three-point function, or angular bispectrum, of the small-angular-scale ($1' \lesssim \theta \lesssim 10'$, $1000 \lesssim l \lesssim 10{,}000$) millimeter-wave sky. A bispectrum signal model that includes contributions from the thermal SZ effect, the clustered cosmic infrared background, and the spatially uncorrelated (or Poisson) point-source signal in each of the three bands provides a reasonable fit to the data. The tSZ bispectrum is detected at $>10\sigma$, the Poisson point-source component is detected in each band individually at $\sim$5 to $\sim 11\sigma$, and the clustered CIB bispectrum is detected at $>5\sigma$. This is the first detection of the clustered CIB bispectrum.

We have compared the measured Poisson point-source bispectrum in each band to predictions from source models. We find that no combination of models of radio-loud and dusty, radio-quiet sources can reproduce the measured Poisson bispectrum amplitudes, implying that bispectrum measurements can provide interesting new constraints on source models.

Applying the methods originally presented in Bhattacharya et al. (2012, B12), we have used the measurement of $B_{\rm tSZ}$, the amplitude of the tSZ bispectrum to constrain σ_8, the normalization of the matter power spectrum, and to predict $A_{\rm tSZ}$, the amplitude of the tSZ power spectrum. The constraint on σ_8 using just SPT bispectrum data and a prior on Ω_b is $\sigma_8 = 0.786 \pm 0.031$. This constraint is competitive with, and statistically consistent with, other recent measurements. Our bispectrum-derived prediction for $A_{\rm tSZ}$, combined with the power spectrum results of Reichardt et al. (2012, R12), results in some evidence for a non-zero kinematic SZ power spectrum, with $A_{\rm kSZ} = 2.9 \pm 1.6\,\mu$K^2, or $A_{\rm kSZ} = 2.6 \pm 1.8\,\mu$K^2 if the $A_{\rm kSZ} > 0$ prior is removed.

In addition to constraining cosmology and models of source emission, these measurements of the small-scale, secondary-anisotropy- and foreground-dominated bispectra provide valuable constraints on potential contamination to measurements of the primordial CMB bispectrum on larger scales, such as those expected soon from the *Planck* team.

We thank Blake Sherwin for providing the ACT l-space filter function, and we thank an anonymous referee for helpful comments. T. Crawford and R. Keisler thank the University of Texas Department of Astronomy and the Texas Cosmology Center, where much of this work was done, for their hospitality.

The SPT is supported by the National Science Foundation through grant ANT-0638937, with partial support provided by NSF grant PHY-1125897, the Kavli Foundation, and the Gordon and Betty Moore Foundation. The McGill group acknowledges funding from the National Sciences and Engineering Research Council of Canada, Canada Research Chairs program, and the Canadian Institute for Advanced Research. Work at Harvard is supported by grant AST-1009012. S. Bhattacharya acknowledges support from NSF grant AST- 1009811, R. Keisler from NASA Hubble Fellowship grant HF-51275.01, B. Benson from a KICP Fellowship, M. Dobbs from an Alfred P. Sloan Research Fellowship, O. Zahn from a BCCP fellowship, and L. Knox and M. Millea from NSF grant 0709498.

Some of the results in this paper have been derived using the HEALPix (Górski et al. 2005) package. This research used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and resources of the University of Chicago Computing Cooperative (UC3), supported in part by the Open Science Grid, NSF grant PHY-1148698. We acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science.

APPENDIX

In this work, we predict the tSZ power spectrum amplitude from the measured tSZ bispectrum amplitude, and we combine that prediction with the R12 measurement of the total SZ power spectrum. Our analysis includes the measurement uncertainties of the power spectrum and the bispectrum, but we assume that the covariance between the two observables is negligible. Here we compute the covariance between the power spectrum and the bispectrum and show that our assumption is justified.

The covariance between the bispectrum (B) and the power spectrum (C) is given by (Kayo et al. 2013)

$$\mathrm{Cov}[C(l_4)B(l_1,l_2,l_3)] = \delta_{l_4 l_1}\frac{4\pi}{\Omega_s l_1 \Delta l_1} C(l_4)B(l_1,l_2,l_3) + 2\,\mathrm{perms.} + \frac{1}{\Omega_s}\int \frac{d\phi}{2\pi} T_5(l_4,-l_4,l_1,l_2,l_3;\phi), \quad \text{(A1)}$$

where Ω_s is the survey area, T_5 is the tSZ five-point function, and ϕ is the angle between l_4 and l_1. Since we use the combined measurements of the power spectrum and the bispectrum at a similar l range, we compute the covariance for the case when $l_4 = l_1$. We do not include the correlated sample variance term from Kayo et al. (2013), because we assume here that the correlation between the sample variance of the power spectrum and the bispectrum is negligible.

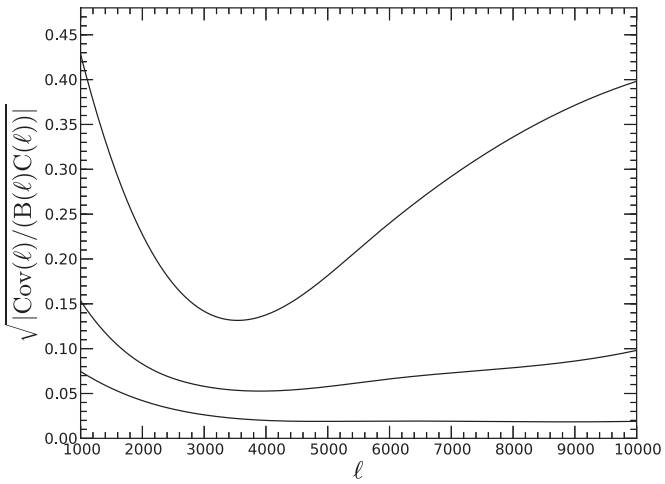


Figure 4. Square root of the fractional covariance between the tSZ bispectrum amplitude B and the tSZ power spectrum amplitude C as a function of l for the different mass cuts. The three-dimensional quantities $\mathrm{Cov}[C(l_1)B(l_1, l_2, l_3)]$ and $B(l_1, l_2, l_3)$ are contracted to one dimension as described in the text. From top to bottom, this quantity is plotted for no cluster cut, $M_{200}(\rho_{\rm crit}) \geqslant 8\times10^{14}\ M_\odot\ h^{-1}$ clusters cut, and $M_{200}(\rho_{\rm crit}) \geqslant 3\times10^{14}\ M_\odot\ h^{-1}$ clusters cut.

We calculate the fractional covariance, $\sqrt{|\mathrm{Cov}(l_{\rm rad})|/|B(l_{\rm rad})C(l_{\rm rad})|}$, where $l_{\rm rad}$ is defined in Equation (29), and $\mathrm{Cov}(CB)$ and B are contracted from 3D to 1D as in Equations (30) and (31). The results are shown in Figure 4 for the three different mass cuts used in the bispectrum estimate (no cut, $M_{200}(\rho_{\rm crit}) > 8\times10^{14}\ M_\odot\ h^{-1}$ clusters cut, and $M_{200}(\rho_{\rm crit}) > 3\times10^{14}\ M_\odot\ h^{-1}$ clusters cut). For a given mass cut, at lower l, the covariance increases while the power spectrum and bispectrum decrease slightly. Hence the fractional covariance increases at lower l. At higher l, the power spectrum and the bispectrum decrease, with the bispectrum decreasing slightly faster than the covariance. This is because the last term in Equation (A1) of the Appendix drops quickly at large l, while the first three terms stay non-zero (as they are the product of the bispectrum and the power spectrum). The net result is that at large l, the ratio increases again. The minimum of the ratio appears at $l \sim 3000$ as both the power spectrum and the bispectrum peak in that range. The covariance decreases sharply with mass cut. This occurs because, with more clusters masked out, the last term of Equation (A1) drops quickly. As shown in Figure 4, the fractional covariance is ≈6% at $l \sim 3000$ for the mass cut $M_{200}(\rho_{\rm crit}) = 8\times10^{14}\ M_\odot\ h^{-1}$. This is small compared to the other sources of uncertainty in the $A_{\rm tSZ}$–$B_{\rm tSZ}$ calculation.

REFERENCES

Acquaviva, V., Bartolo, N., Matarrese, S., & Riotto, A. 2003, NuPhB, 667, 119
Addison, G. E., Dunkley, J., & Bond, J. R. 2013, MNRAS, 436, 1896
Addison, G. E., Dunkley, J., Hajian, A., et al. 2012, ApJ, 752, 120
Allen, S. W., Rapetti, D. A., Schmidt, R. W., et al. 2008, MNRAS, 383, 879
Arnaud, M., Pratt, G. W., Piffaretti, R., et al. 2010, A&A, 517, A92
Benson, B. A., de Haan, T., Dudley, J. P., et al. 2013, ApJ, 763, 147
Béthermin, M., Daddi, E., Magdis, G., et al. 2012, ApJL, 757, L23
Béthermin, M., Dole, H., Lagache, G., Le Borgne, D., & Penin, A. 2011, A&A, 529, A4
Bhattacharya, S., Heitmann, K., White, M., et al. 2011, ApJ, 732, 122
Bhattacharya, S., Nagai, D., Shaw, L., Crawford, T., & Holder, G. P. 2012, ApJ, 760, 5
Butcher, H., & Oemler, A., Jr. 1984, ApJ, 285, 426
Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. 2011, PASP, 123, 568
Coles, P., & Barrow, J. D. 1987, MNRAS, 228, 407
Das, S., Louis, T., Nolta, M. R., et al. 2013, arXiv:1301.1037
Das, S., Marriage, T. A., Ade, Peter A. R., et al. 2011a, ApJ, 729, 62
Das, S., Sherwin, B. D., Aguirre, P., et al. 2011b, PhRvL, 107, 021301
De Zotti, G., Massardi, M., Negrello, M., & Wall, J. 2010, A&ARv, 18, 1
Dudley, J. 2012, PhD thesis, McGill Univ.
Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64
Dunkley, J., Hlozek, R., Sievers, J., et al. 2011, ApJ, 739, 52
Fergusson, J. R., & Shellard, E. P. S. 2009, PhRvD, 80, 043510
Finkbeiner, D. P., Davis, M., & Schlegel, D. J. 1999, ApJ, 524, 867
Fowler, J. W., Acquaviva, V., Ade, P. A. R., et al. 2010, ApJ, 722, 1148
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Gruzinov, A., & Hu, W. 1998, ApJ, 508, 435
Haiman, Z., Mohr, J. J., & Holder, G. P. 2001, ApJ, 553, 545
Hall, N. R., Keisler, R., Knox, L., et al. 2010, ApJ, 718, 632
Hashimoto, Y., Oemler, A., Jr., Lin, H., & Tucker, D. L. 1998, ApJ, 499, 589
Hasselfield, M., Hilton, M., Marriage, T. A., et al. 2013, JCAP, 07, 008
Hauser, M. G., & Dwek, E. 2001, ARA&A, 39, 249
Hill, J. C., & Sherwin, B. D. 2013, PhRvD, 87, 023527
Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19
Hinshaw, G., Spergel, D. N., Verde, L., et al. 2003, ApJS, 148, 135
Hivon, E., Górski, K. M., Netterfield, C. B., et al. 2002, ApJ, 567, 2
Holder, G. 2014, ApJ, 780, 112
Holder, G. P. 2002, ApJ, 580, 36
Hou, Z., Reichardt, C. L., Story, K. T., et al. 2014, ApJ, 782, 74
Hu, W. 2000, ApJ, 529, 12
Kayo, I., Takada, M., & Jain, B. 2013, MNRAS, 429, 344
Keisler, R., Reichardt, C. L., Aird, K. A., et al. 2011, ApJ, 743, 28
Knox, L., Scoccimarro, R., & Dodelson, S. 1998, PhRvL, 81, 2004
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
Komatsu, E., & Seljak, U. 2002, MNRAS, 336, 1256
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Komatsu, E., & Spergel, D. N. 2001, PhRvD, 63, 063002
Lacasa, F. 2012, arXiv:1204.1480
Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16
Lueker, M., Reichardt, C. L., Schaffer, K. K., et al. 2010, ApJ, 719, 1045
Mocanu, L. M., Crawford, T. M., Vieira, J. D., et al. 2013, ApJ, 779, 61
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Nozawa, S., Itoh, N., Kawana, Y., & Kohyama, Y. 2000, ApJ, 536, 31
Ostriker, J. P., & Vishniac, E. T. 1986, ApJL, 306, L51
Plagge, T. J., Marrone, D. P., Abdulla, Z., et al. 2013, ApJ, 770, 112
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A8
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, A&A, 550, A131
Press, W., Flannery, B., Teukolsky, S., & Vetterling, W. 1986, Numerical Recipes: The Art of Scientific Computing (Cambridge: Cambridge Univ. Press)
Reichardt, C. L., Ade, P. A. R., Bock, J. J., et al. 2009, ApJ, 694, 1200
Reichardt, C. L., Shaw, L., Zahn, O., et al. 2012, ApJ, 755, 70
Reichardt, C. L., Stalder, B., Bleem, L. E., et al. 2013, ApJ, 763, 127
Riess, A. G., Macri, L., Casertano, S., et al. 2011, ApJ, 730, 119
Santos, M. G., Balbi, A., Borrill, J., et al. 2002, PhRvL, 88, 241302
Sayers, J., Czakon, N. G., Mantz, A., et al. 2013, ApJ, 768, 177
Schaffer, K. K., Crawford, T. M., Aird, K. A., et al. 2011, ApJ, 743, 90
Sehgal, N., Bode, P., Das, S., et al. 2010, ApJ, 709, 920
Shaw, L. D., Nagai, D., Bhattacharya, S., & Lau, E. T. 2010, ApJ, 725, 1452
Shaw, L. D., Rudd, D. H., & Nagai, D. 2012, ApJ, 756, 15
Shaw, L. D., Zahn, O., Holder, G. P., & Doré, O. 2009, ApJ, 702, 368
Shirokoff, E., Reichardt, C. L., Shaw, L., et al. 2011, ApJ, 736, 61
Sievers, J. L., Hlozek, R. A., Nolta, M. R., et al. 2013, JCAP, 10, 060
Smith, S., Rocha, G., Challinor, A., et al. 2004, MNRAS, 352, 887
Stanek, R., Rasia, E., Evrard, A. E., Pearce, F., & Gazzola, L. 2010, ApJ, 715, 1508
Story, K. T., et al. 2013, ApJ, 779, 86
Sunyaev, R., & Zel'dovich, Y. 1980, ARA&A, 18, 537
Sunyaev, R. A., & Zel'dovich, Y. B. 1970, CoASP, 2, 66
Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
Tucci, M., Toffolatti, L., de Zotti, G., & Martínez-González, E. 2011, A&A, 533, A57
van Engelen, A., Keisler, R., Zahn, O., et al. 2012, ApJ, 756, 142
Vanderlinde, K., Crawford, T. M., de Haan, T., et al. 2010, ApJ, 722, 1180
Vieira, J. D., Crawford, T. M., Switzer, E. R., et al. 2010, ApJ, 719, 763
Vieira, J. D., Marrone, D. P., Chapman, S. C., et al. 2013, Natur, 495, 344
Viero, M. P., Wang, L., Zemcov, M., et al. 2013, ApJ, 772, 77
Wang, L., & Steinhardt, P. J. 1998, ApJ, 508, 483
Williamson, R., Benson, B. A., High, F. W., et al. 2011, ApJ, 738, 139
Wilson, M. J., Sherwin, B. D., Hill, J. C., et al. 2012, PhRvD, 86, 122005
Yadav, A. P. S., Komatsu, E., & Wandelt, B. D. 2007, ApJ, 664, 680
Yadav, A. P. S., & Wandelt, B. D. 2008, PhRvL, 100, 181301
Yadav, A. P. S., & Wandelt, B. D. 2010, AdAst, 2010, 71
Zahn, O., Reichardt, C. L., Shaw, L., et al. 2012, ApJ, 756, 65